

French science after Chirac

Presidential elections next spring are set to lend fresh impetus to research reform in France.

"Time for a French revolution", "A chance for change in France", "France's research base needs a shaking", "A squandered opportunity", "France: too much analysis leads to paralysis?". These headlines from *Nature* have punctuated more than a decade of successive opportunities, false starts and failures in the reform of France's research system.

Is French science, and indeed France itself, now ready for a long-overdue renewal? The country is plagued with high unemployment, strained public finances and a faltering economy. Its overall scientific effort has stagnated for more than a decade. Yet meaningful reform of either French society or its research system have proved elusive, with a vicious spiral of inept attempted change and spirited, usually successful, resistance making reform seem almost impossible.

To believe that France is somehow unreformable would be simplistic, however, and there is currently good reason to be optimistic. Next year's presidential election offers the hope of a fresh start. The favourites for the election, socialist Ségolène Royal and conservative Nicolas Sarkozy, have each built their campaigns on their youth and the need for radical change. That matches the mood of the country, which seems ready to embrace a break with a passing political generation that seems ineffective and out of touch.

In science too there are signs of change. Jacques Chirac's two terms of office have seen an overall decline in French science, and an unprecedented level of disharmony between scientists and the government. The administration has shown a disregard for science and a penchant for attempting to implement change without consultation, and has made deep cuts in science budgets and jobs. But given the likelihood of an improved climate after the presidential election, there are signs that France may be ready to modernize its research system.

To its credit, the Chirac administration has created some of the tools that could enable change. In particular, it has set up a National Research Agency (ANR) to distribute research funds on the basis of grant proposals, creating greater competition than previously existed, when funds were largely distributed to laboratories directly.

Researchers seem to have now accepted the principle of the shift to this more Anglo-Saxon dual system of labs and competitive funds. But they remain deeply sceptical of the agency itself, which to many looks like an arm of government destined to centralize control over

research, lacking the scientific autonomy and authority of, say, the US National Science Foundation. This scepticism has some basis, and given the government's general neglect of science, it is unsurprising that the scientific community has not fully accepted the new agency.

The principle behind the ANR is sound, however. Corresponding shifts seem also to be under way at the CNRS, Europe's largest basic research agency, which early next year will present a series of reforms, probably including higher salaries for researchers who score better on international performance indicators, and a focusing of resources on fewer areas of excellence. Such changes are needed at an agency notorious for being inflexible and bureaucratic.

CNRS management has already created controversy this month by announcing that it will rein back support for the life sciences, except for neurosciences and systems biology, claiming that researchers' performance is not up to scratch in this sector, which consumes a third of the agency's budget. It is promising that the CNRS seems willing to make tougher choices with its own budget, to support excellence while maintaining a balance of disciplines.

This should also, however, provoke a wider debate on the country's overall research priorities. It is true that the impact of France's life sciences in international publications is below that of other fields, such as mathematics and physics. But this also reflects the fact that France, with its focus on large programmes in aerospace and nuclear research, has given life sciences less support than has Britain, for example. If French life sciences are less than competitive, it is partly because they have never been given the means to become so.

Despite a tradition of resistance to change, many in the French research community yearn for a more modern, more competitive and less bureaucratic research system that would give the brightest young researchers the autonomy they need to take French science forward. The next president of France, and the government that he or she appoints, will have a crucial role in making research a genuine priority, and working with, rather than against, the scientific community. French science can ill afford yet another false start. ■

"Many in the French research community yearn for a more modern, more competitive and less bureaucratic research system."

Smiles all round

Can research buy you happiness?

When physicist Paul Rothemund of the California Institute of Technology created smiley faces by the billion out of self-assembling pieces of DNA (*Nature* **440**, 297–302; 2006) — an achievement celebrated on the cover of this journal — it

gave his friend and colleague Erik Winfree a nice opportunity to joke that the work represented "the most concentrated happiness ever created". But in general, although happiness may often be a by-product of scientific research — most notably for the researchers themselves — its direct production, let alone its concentration, is not.

Yet happiness is not a purely subjective phenomenon, immune to scientific analysis. As we report on page 418, it is a focus of interest on the parts of both researchers and policy-makers. A growing number of people are calling for happiness, and measures to enhance it, to



be given greater weight in government policy. Richard Layard of the London School of Economics and Political Science, who also has a seat in Britain's House of Lords, explicitly points to what he sees as a "new science" of happiness as the basis for promoting radical shifts in political goals.

Unfortunately, this new science has yet to provide a compelling account of how happiness is created. Instead, for obvious methodological reasons, it concentrates on what it correlates with. But it is not clear that changing those correlates by dictat would necessarily produce the desired effect. People may be happy spending time with their children, but forcing parents to spend more time this way would not necessarily overjoy everyone involved. Expressing gratitude makes people happier; a politeness police, probably, would not.

There is further debate as to whether trying to do something about

people's happiness is feasible in principle. Some researchers favour the idea that people have a 'hedonic setpoint' that stays remarkably constant in the face of bouquets and brickbats. Such a setpoint need not in principle be unalterable, but its alteration might require an approach with a pharmacological component, raising the problem that one of the things we value about happiness is its authenticity. Another is its autonomy. Governments may guarantee citizens freedom in their pursuit of happiness, but we bridle against the idea of its ever being enforced.

But it is reasonable to hope that through a better understanding of happiness, we may eventually improve the chance of people attaining it. In the meantime, at least the wonders of science and nature, from smiley-face DNA to footwear for penguins (see page 412), may provide some happiness simply in and of themselves. ■

A fresh start for Japan

Japan has a new prime minister, a new science adviser, and a chance to change its science policy.

For years Japan has been trying to reform the way it does science. Since 2001, almost all research institutions, including the entire university system, have been given far greater autonomy in an attempt to make them more competitive with each other and with rivals abroad. Generous funding has also been ploughed into strategic research projects.

The Protein 3000 project, for example, has sought to find the structures of 3,000 proteins in five years, ending next March. These structures will, it is hoped, prove useful in understanding larger proteins that are difficult to analyse directly because of their size. A project called BioBank has also assembled a huge collection of disease-related tissue samples: from June 2003 until this October, it took samples from 162,545 patients — a total of 231,049 disease-related samples, because some patients have more than one disease — on its way to meeting its goal of taking samples from 300,000 patients by March 2008. The samples will be used to search for genetic factors behind 47 diseases, including lung and stomach cancer, diabetes and stroke.

But with the institutional reforms well under way and some of these projects set to wind down, the time is now ripe for Japan to revisit its science policy. In September, Shinzo Abe took over as prime minister and pledged to do just that, announcing plans for Innovation 25, a long-term review of Japan's science and technology needs until 2025. He also took the unprecedented step of appointing a scientist, Kiyoshi Kurokawa, as a special adviser to the cabinet. It is commendable that Abe is taking science policy seriously.

There are a couple of recurrent problems in Japanese science that the new government needs to address. One is an academic system that, notwithstanding the recent reforms, allows elite professors to sit at the top of a rigid hierarchy while junior colleagues patiently await their chance to move up the ladder. More changes are needed to shake up this system.

A second problem concerns the conception and design of major

projects such as Protein 3000 and BioBank. These projects have been set up with explicit numerical goals that the public can readily understand. Despite being on target to achieve these stated goals, the government now wonders whether the projects actually best serve Japan's scientific interest. A probable redirection of government support could mean that some of the significant resources that have been invested in these projects — such as the costly machines used to produce protein structures or the samples donated by patients — will go to waste. More meticulous planning and greater consultation with scientists at the planning stage could arguably have assured better use of project funding.

This last issue reflects, in part, project management by civil servants who rotate rapidly through positions every two or three years. They are forced to make important decisions on science policy but are rarely around to be held accountable for outcomes. Japan needs to move more responsibility for the nuts and bolts of such project planning towards advisory boards of scientists who have the necessary expertise and can, at least in theory, be held accountable for the advice that they offer.

One bright spot for science in the new prime minister's record so far is his determination to improve relations with influential neighbours such as South Korea and China. This could — and should — open the way for greater regional collaboration in science and technology.

Kurokawa, meanwhile, has the contacts and experience to lend valuable impetus to Japanese science policy. As a former president of the Science Council of Japan, which represents 790,000 Japanese scientists, he has not been afraid to speak out on scientific issues in the past, and has a broad grasp of the strengths and weaknesses of Japanese science.

The new science adviser should try to find ways of getting more scientists, from both inside and outside Japan, involved in contributing to the development of science policy in Tokyo. With the backing of the prime minister, Kurokawa is well placed to make a difference, and to build on earlier reforms to further increase Japan's scientific productivity. ■

"The new science adviser should try to find ways of getting more scientists involved in the development of science policy."

RESEARCH HIGHLIGHTS

Disruptive influence

Nature Neurosci. doi:10.1038/nn1801 (2006)

Cannabis may work its disruptive effects on memory by interfering with the synchronous firing of brain cells, researchers have found.

György Buzsáki of Rutgers, The State University of New Jersey in Newark, and his colleagues examined the timing and frequency of neuronal signals in animals injected with the active component of marijuana. They focused on the hippocampus, a brain area involved in the formation of memories.

Firing rates of hippocampal neurons were marginally affected by the cannabinoids, but the timing of the signals, which normally show a degree of synchrony, was disrupted. The extent of this disruption correlated with the animals' memory impairment.



T. MARECHAL/ALAMY

NANOTECHNOLOGY

What a contrast

Nature Mater. doi:10.1038/nmat1775 (2006)

Nanoparticles of iron-cobalt alloy are tiny magnets that could enhance the contrast of magnetic-resonance imaging (MRI) — used for medical scans — when injected into the body. But they 'rust' easily and are toxic, so have been of little use for such applications.

Now Hongjie Dai and his colleagues at Stanford University, California, have figured out how to wrap these particles in a thin, protective shell of carbon. They make the coats by exposing the alloy to carbon vapour, then attach phospholipids to make the particles water-soluble. These nanoparticles gave greater MRI contrast enhancement than commercial alternatives *in vivo*, without any apparent toxicity at the doses required.

BEHAVIOURAL GENETICS

Girls shove, boys box

Nature Neurosci. doi:10.1038/nn1809 (2006)

Differences between the fighting styles of female and male fruitflies are determined by a gene known as *fruitless*, researchers have found. The results imply that aggressive behaviours are hardwired into the nervous system of *Drosophila* flies.

Males with a female variant of *fruitless* fought female-style — with head-butts and shoving — whereas females with a male variant fought male-style, lunging and boxing. Modified

males also failed to establish dominance.

The *fruitless* gene, which also dictates male courtship behaviour, is transcribed differently in male and female flies. Barry Dickson of the Research Institute of Molecular Pathology in Vienna, Austria, and his colleagues made their observations by engineering flies to produce transcripts typical of flies of the opposite sex.

CLIMATE SCIENCE

How low did it go?

Earth Planet. Sci. Lett. doi:10.1016/j.epsl.2006.09.033 (2006)

A uniquely well-preserved record of past sea levels in the Gulf of Lions in the Mediterranean has been used to constrain the sea's lowest points over the past 434,000 years.

Marina Rabineau of the National Centre of Scientific Research in Plouzané, France, and her colleagues used high-resolution seismic profiles and shallow drilling cores to determine the position of palaeoshorelines at the peaks of the last five glacial periods (in

graphic below, red dots indicate new data, grey areas the range of past estimates).

After correcting for the effects of tectonic movement and sediment build-up, they conclude that sea levels were around 100 metres lower than today during the three most recent glacial periods, and around 150 metres lower during the earlier two. The difference points to more extreme glaciations in earlier cycles, which may be associated with the 400,000-year cycle in the eccentricity of Earth's orbit.

CANCER BIOLOGY

Addicted to genes

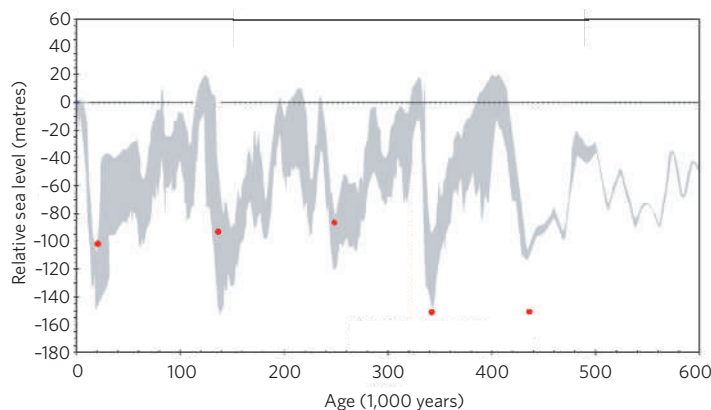
Cancer Cell 10, 425–435 (2006)

Tumour cells' apparent 'addiction' to the mutated genes that first turned them cancerous is examined in a new study from Jeffrey Settleman of Harvard Medical School in Charlestown, Massachusetts, and his team.

Certain types of cancerous cell die when their oncogenes are knocked out — a phenomenon that has been dubbed 'oncogene

addiction'. The researchers propose that this happens not because the cells are dependent on the gene, but because switching off the gene sets up an imbalance between signals that promote cell survival and those that promote cell death, or apoptosis.

They tested the idea in a number of cell lines, finding that following oncogene inactivation, certain proapoptotic signals do persist for longer than prosurvival signals.



PHYSICS

All together now

Phys. Rev. Lett. **97**, 187402 (2006)

Physicists have gathered the most direct evidence yet that particles known as excitons can form a 'condensate'.

Excitons form when a negatively charged electron pairs with a positively charged 'hole' in a semiconductor. Theorists have predicted that a group of excitons can occupy a single quantum state and behave like one giant particle. The state is predicted to resemble one previously seen for ultracold atoms, known as a Bose–Einstein condensate.

Leonid Butov of the University of California, San Diego, and his team used a laser to create a ring of excitons in a semiconductor wafer. They saw that, when the excitons were cooled below a few Kelvin, their properties synchronized over micrometre scales. This spontaneous coherence indicates that a condensate has formed, but it is not yet clear whether the excitons' behaviour will match that of atoms in a conventional Bose–Einstein condensate.

MOSQUITO-BORNE DISEASES

Senior citizens

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0604875103 (2006)

A new technique to estimate a mosquito's age should aid efforts to understand the spread of diseases such as malaria and dengue.

Mosquitoes can only transmit such diseases once they are old enough to have incubated infective parasites — typically after 12 days of adult life. Existing methods to estimate a mosquito's age involve sampling chemicals from the surface of the insect's legs, but this only allows estimation of ages up to 15 days and they can live for longer than this.

Scott O'Neill of the University of Queensland in Brisbane, Australia, and his



colleagues identified a suite of genes in the mosquito *Aedes aegypti* that alter the genes' expression patterns over time. Estimating insect ages by profiling these genes works for more senior insects and is more accurate than previous techniques.

EVOLUTION

Which came first?

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0608762103 (2006)

The genetic code predates the enzymes that act as its stewards, argue researchers at Yale University in New Haven, Connecticut.

Dieter Söll and his colleagues studied enzymes known as aminoacyl-tRNA synthetases. These help to translate the genetic code into protein sequences by attaching amino acids to transfer RNAs (tRNAs).

They found that enzymes from two other microbes could interact with a tRNA isolated from the archaeal microorganism *Methanocaldococcus jannaschii*, charging it up with different amino acids.

This suggests that the enzymes evolved in the different organisms to accommodate a tRNA whose role in translation was already fixed. The tRNA may have emerged in the 'RNA world' that some people think predated the modern world of DNA and proteins.

GENETIC ENGINEERING

Pig lets vector stay

Proc. Natl Acad. Sci. USA **103**, 17672–17677 (2006)

A novel vector for delivering foreign DNA into mammalian cells has been tested in pigs.

Most approaches to gene therapy or genetic engineering use custom-built DNA vectors based on viral genomes, which insert themselves into the host cells' DNA. But if the DNA inserts in the wrong place, it may not function properly and can cause mutations.

Marialuisa Lavitrano of the University of Milano-Bicocca, Italy, and her colleagues have developed a vector that can stay in cells without integrating into the host's genome, and which replicates independently. They added this 'nonviral episomal vector' to pig sperm, which they then used to fertilize eggs. Of the 18 resulting fetuses, 12 carried the vector in all tissues tested.

MOLECULAR BIOLOGY

Over the fold

Cell **127**, 803–815 (2006)

The lung disease cystic fibrosis occurs when mutations prevent a membrane channel protein from folding up correctly. Tweaking a component of the folding machinery can help, show John Yates and William Balch of the Scripps Research Institute in La Jolla, California, and their colleagues.

The team pinpointed the stage at which folding goes awry by comparing the protein interactions of the mutant transmembrane regulator CFTR with those of its normal counterpart.

They found that partially blocking a helper protein, known as a co-chaperone, that is involved in stalling folding helped the mutant protein to form properly. Exactly how it does so is not yet clear, but the approach might also work for other diseases that involve protein misfolding, such as Alzheimer's.

JOURNAL CLUB

Ralph Lewin
Scripps Institution of
Oceanography, La Jolla,
California, USA

A marine biologist sees the potential of cyanobacteria, and the benefits of their renaming.

Let's start with a false syllogism: bacteria are prokaryotes, blue-green algae are prokaryotes, and therefore blue-green algae are bacteria. All other algae are

eukaryotes and so, the argument went, we should reclassify the Cyanophyta as cyanobacteria.

I was never in favour of this renaming, but it may have been good for funding. I've heard that grant applications for research on bacteria have better chances of success than those for research on blue-green algae.

And these oft-neglected organisms have a lot to offer. A recent paper on *Lyngbya majuscula* from Bill Gerwick, now at the Scripps Institution

of Oceanography in La Jolla, California, and his colleagues (B. Han *et al.* *J. Nat. Prod.* **69**, 572–575; 2006), for example, reveals some interesting new compounds.

L. majuscula grows on warm seashores as tufts, which, when they come loose and float away, can stick to swimmers' skin and cause a rash — known as swimmers' itch or seaweed dermatitis.

Gerwick and his team extracted from dried *L. majuscula* two compounds that may explain its irritant effect. The compounds,

aurilide B and aurilide C, are hugely complicated ring-shaped molecules that resemble a toxin previously isolated from sea slugs.

In tissue culture assays, the compounds proved toxic to human and mouse cancer cells. Such natural products can act as starting points for pharmaceutical chemists.

Gerwick's paper refers to *L. majuscula* as a cyanobacterium in its title and as an alga elsewhere in its text, but what's important is the science, not the names.

SPECIAL REPORT

Anti-evolutionists raise their profile in Europe

The teaching of alternative theories to evolution in schools is not just an issue in the United States. **Almut Graebisch** and **Quirin Schiermeier** assess whether creationism is threatening science in Europe.

Being a trained biologist doesn't stop Maciej Giertych from insisting that evolution is a falsified hypothesis¹. The 70-year-old Polish member of the European parliament, who has a PhD in tree physiology, also wants to spread the word. In October, he organized a workshop for parliamentarians entitled "Teaching evolution theory in Europe: is your child being indoctrinated in the classroom?"

Although the teaching of evolution has become a highly politicized and hotly discussed matter in the United States, such moves are rare in Europe, and Giertych's activities have so far met with little response in Strasbourg or Brussels. But a number of similar incidents over the past couple of years, in various countries, are raising fears among the scientific community that creationism may be on the rise in Europe.

Last month, for example, it emerged that creationism is being taught at two schools in the German state of Hesse. The incident, albeit minor, has provoked debate in the country. The Christian view of creation should at least be discussed in science classes, argues Karin Wolf, Hesse's Christian Democrat education minister. But the Association of German Biologists warns of the dangers of blurring the division between science and religion.

And in Britain in September, the prominent creationist group Truth in Science sent information packs to every UK secondary school. The material suggests intelligent design should be taught as an alternative to the theory of evolution, although the UK government's education department was quick to say that it does not endorse its use in science classes.

In response, a group called the British Centre for Science Education has been formed to campaign against the teaching of creationism in schools. Meanwhile, British school leavers' knowledge about evolution is considered so poor, and creationist ideas so widespread, that

the universities of Leeds and Leicester are planning to introduce remedial courses next year for first-year science students.

Steve Jones, a geneticist at University College London who has lectured widely about evolution, is one of those concerned by the growing influence of creationist groups. "I have talked about evolution in front of more than 100,000 British schoolchildren in the past 20 years — during most of that time I was never asked questions about creationism," he says. "But in the past couple of years, wherever I go I am asked about it." He ascribes the change largely to the activities of groups such as Truth in Science.

But perhaps the most blatant attempt to ban evolution from the classrooms occurred in Italy in 2004. Letizia Moratti, then education minister, caused a public outcry when she removed the theory of evolution from the curricula of Italy's middle schools² on the grounds that teaching Darwin's theory of evolution can instil a materialist view of life in young minds.

Following widespread protest, the education ministry partially reintroduced darwinism into school courses. A recent study by *Observa Science in Society*, a Vicenza-based body that promotes informed debate on scientific issues, shows that only 11% of Italians support the exclusion of darwinism from curricula. But almost two-thirds would prefer lessons to cover both evolutionary theory and the creationist view. "Italy is no longer a completely secular country," says Telmo Pievani, a philosopher of science at the University of Milan II in Italy. "We are facing a dramatic and worrying cultural and political regression."

In Russia, meanwhile, creationist societies are receiving strong support from the Protestant minority. Besides translating the writings of European and US creationists, Russian groups conduct their own 'creation research'. In Moscow, for example, the ARCTUR Research Geological Lab is looking for geological and

"Creationism is a major issue in Turkish politics; the debate is much more tense than in the United States."



geochemical proof of creationism. The society collaborates with creationists in the West and promotes its findings in several Russian and English-language creationist journals.

Such examples illustrate the complexity of the issue in Europe compared with the United States. Whereas the US drive towards creationism comes mainly from Protestant fundamentalist groups, the European movement has diverse roots. "There is an aggressive anti-darwinism inspired by radical Islamic minorities in immigrant communities in Britain and France; there is a Catholic creationism growing in Poland; there is Protestant creationism in some schools in England," says Pievani.

The Koran is less clear than the Bible on divine creation. But that does not mean Islam accepts evolution, and the influence of Islamic creationist groups in countries such as Britain and France is increasing. The movement is by far the strongest in Turkey, however, which is in negotiations to join the European Union. The main Muslim creationist organization, the Turkish Bilim ve Araştırma Vakfı, distributes creationist literature in Turkey and elsewhere that often consists of material translated from Christian fundamentalist groups, particularly the Institute for Creation Research, based in



RELIGION AND SCIENCE IN FOCUS

Find all of *Nature's* stories on this contentious topic online.
www.nature.com/news/infocus



The teaching of evolution theory is under threat in some European countries.

North Santee, California. Prominent US creationists are also frequently invited by the group to give talks.

Jones has just returned from the Istanbul book fair, where he says many creationist publications were on sale, and proving extremely popular. "Creationism is a major issue in Turkish politics; the debate is much more tense than in the United States," he says. "All biology textbooks now used in schools are creationist in tone."

There is debate over the size of the threat posed to science in Europe by the various creationist movements. The creationists' main goal is to have their views included in school curricula. But, unlike in the United States, many pupils in European schools receive religious education anyway and are therefore familiar with the theme of divine creation. Some think this may steal the thunder of the creationist movement in Europe — why force creationism into science classes if it is already taught in religious education?

Moreover, Europeans do seem to be more enlightened than Americans when it comes to evolution. According to a 2005 US study³, just 40% of Americans accepted the theory of evolution, down from 45% in 1985. In Europe, that figure has increased, from 65% in 1992 to 70% in 2005, although numbers vary widely across the region⁴.

"It annoys and depresses me that intelligent students persist in holding irrational views."

Others warn that scientists can't afford to be complacent. "The anti-evolution movement does undermine public understanding of science," argues Ulrich Kutschera, an evolutionary biologist at the University of Kassel in Germany and vice-president of the Association of German Biologists. "In Germany and other European countries, anti-evolutionists with different religious backgrounds promote their ideology via colourful web pages

that are appealing to students and people without a scientific background. Perception of evolutionary biology, in particular, is seriously undermined by these activities." He suggests that biologists should devote more time to counteract-

ing its spread by explaining their theories to the public.

Jones says that, despite his dislike of creationism — "it annoys and depresses me that intelligent students persist in holding irrational views" — he doesn't think that such arguments are set to undermine science in countries such as Britain. "But I am not so optimistic about Turkey."

1. Giertych, M. *Nature* **444**, 265 (2006).
2. *Nature* **428**, 595 (2004).
3. Miller, J. D. et al. *Science* **313**, 765–766 (2006).
4. Special Eurobarometer Europeans, Science & Technology http://ec.europa.eu/public_opinion/archives/ebs/ebs_224_report_en.pdf (Eur. Comm., 2005).

Q&A PETER KOREVAAR

Peter Korevaar is head of the physics and cosmology working group of Germany's Studiengemeinschaft Wort und Wissen, one of the largest creationist groups in Europe. He holds a PhD in astrophysics and now works at IBM in Mannheim. Quirin Schiermeier asks him about his group's aims.

What are your main goals?

We are a Protestant group. We want to do accurate and honest scientific work under the premise that God has created the world. Scientific naturalism as we know it doesn't allow for a creator who can interfere with the physical world. Evolution should be taught in schools, and creation discussed along with it.

How would you describe your relationship with scientists?

You don't have to agree on everything to do good, accurate science together with [non-creationist] scientists. We use the same methods as other scientists,

namely falsifying and verifying hypotheses. We don't want to put anyone down.

We would very much like to have an open discussion with evolutionary biologists about the issues at stake. But we feel constantly misunderstood. Scientists — and the media — always say we are dilettantes, Christian fundamentalists. This is mean. 'Fundamentalism' is immediately associated with Islamic fundamentalism: read terrorism. Fighting against these prejudices



is extremely hard.

What about evolution?

Microevolution, the adaptation of species to their environment, is an observed scientific fact, which we of course do not deny.

But macroevolution, the gradual process of development of new species, is a mere conclusion, there's no observational evidence for that.

How would you compare your group with creationists

in the United States?

We are aware of other creationist groups in Europe and the United States. But we don't collaborate too much with any of them. The US debate is more aggressive, there is more foul play from both sides. This is not helpful.

Do you advocate intelligent design?

There's an open question about how the many complex structures observed in the Universe came into being. Intelligent design gives an alternative answer to this question. We can subscribe to most of its arguments.

Tragedy increases woes of US ice fleet

The main US ship for Arctic research is patching together plans for next year's cruises under the shadow of safety concerns, after its marine science officer and a crewman died while diving under the ice.

The US Coast Guard icebreaker *Healy* is scheduled to begin operations in April, as part of International Polar Year, in which hundreds of scientists from at least 60 countries will launch projects at the poles. But US agencies are experiencing considerable challenges, including replacing two ageing icebreakers, securing a crucial helicopter service, and training icebreaker crews for scientific missions.

And in a telephone conference on 16 November, Coast Guard officers told US research agencies that divers may not be available for the 200 days of scientific operations that the *Healy* is scheduled to complete. A Navy-certified dive team typically supports researchers aboard the 128-metre icebreaker, helping with equipment and operations in the water.

On 17 August, while stopped in pack ice about 800 kilometres north of Barrow, Alaska, marine science officer Jessica Hill and a mate, Steven Duque, died while on a short dive under the ice. Hill, a marine biologist who was also the ship's dive officer, was conducting a 'familiarization' dive for the less-experienced Duque.

When the divers ran into difficulties, crew members pulled them out of the water via life-

line ropes. Resuscitation was unsuccessful and the divers were pronounced dead. The *Healy* steamed to Alaska, where Coast Guard officials relieved the captain, suspended all dives, and cancelled the ship's two remaining research cruises for 2006. A month-long cruise led by scientists at the University of New Hampshire in Durham, who are mapping the continental shelf to redraw the US economic zone, was delayed until next September. And a cruise to test an under-ice vehicle operated by researchers from Woods Hole Oceanographic Institution in Massachusetts was cancelled. The team will deploy the vehicle next summer without the benefit of a full test run.

"We were disappointed," says Simon Stephenson, director of the Arctic science division of the US National Science Foundation (NSF). "But this was a safety issue; there was concern about the crew functioning properly."

Coast Guard investigations into the deaths are due to be completed by February 2007. Earlier this month, military officials released autopsy reports to the families in Florida. Hill, 30, died of asphyxiation, with lung trauma from a likely unconscious ascent from about 60 metres down. The report says that both divers' air tanks were empty. "Nothing will bring her back, but I would like to know what happened," says retired microbiologist William Hill, the woman's father.

The tragedy adds to difficulties already being experienced by the US fleet of four icebreakers. In September, the Government Accountability Office reported on the inadequacies of the ageing ships, and a National Academy of Sciences panel called for the replacement of two of them (*Polar Sea* and *Polar Star*).

The *Healy* is the newest of the fleet, commissioned in 2000, and the only one primarily devoted to scientific missions. But there have been difficulties with the *Healy*'s Coast Guard crew having sufficient technical experience to support researchers on board. In recent years, the NSF has had to contract out various services, including the handling of ocean-floor cores (to Oregon State University) and key depth and temperature data (to Scripps Institution of Oceanography). Last summer the agency also had to pay for a contract helicopter service for some cruises, doubling costs to US\$600,000, because Coast Guard helicopter engines were being replaced.

Stephenson is now aiming for a more coordinated approach. Next year, he says, the NSF will issue a request for proposals in which one institution would provide scientific support personnel for the *Healy*. The contract, worth around \$1 million, is to be in place for 2008. But Stephenson is still aiming to fill a month-long gap in June, when no team is scheduled to use the ship. "It's ironic this is happening [during International Polar Year]," he admits. "But I predict there will be too much demand in 2008."

Rex Dalton

P. DANNER/US COAST GUARD/AP

Italian government eases in radical reforms

The tiny, elegant and undeniably ancient Rita Levi-Montalcini — Nobel laureate and, at 97, still active in her role as senator-for-life — thrust the plight of Italian research into the headlines last week. She pledged to vote against the new government's 2007 budget if last-minute cutbacks for science were not reversed.

That small battle was won, and the bill passed through the first of the Italian parliament's two chambers on 18 November with a final budget for universities and research only slightly lower than that for this year.

There is still much fighting to be

done, but when the stringent 2007 budget — designed to reduce Italy's soaring deficit — is finally signed off in December, scientists are likely to see a significant increase in research project money, a decrease in funding for infrastructure, and a smaller-than-anticipated increase in new positions for young academics.

This may disappoint those who had higher hopes for the avowedly research-friendly centre-left coalition government that took office in the summer. But behind the headlines, more fundamental changes to the research system

have quietly taken place. Two key decrees were approved last week, which in the long run may serve the scientific community better than immediate cash injections.

One decree creates an evaluation agency for universities and research, which should be established by the spring. Its first task will be to broadly grade universities on their performance



Defending Italian science: Rita Levi-Montalcini

in teaching and research, producing three or four groupings that the ministry for universities and research will eventually use to allocate funds. "The agency's aim is to gradually improve, not to judge," says Luciano Modica, the

research ministry's undersecretary.

Modica hopes the agency will also help to solve other problems

L. BRUNO/AP



The icebreaker *Healy* is primarily used for scientific research.



CLIMATE CONFERENCE IN NAIROBI

Find out what happened at last week's convention on the Kyoto Protocol.

www.nature.com/news

Britain plans tough limits to curb emissions

The UK government clearly aimed to make a statement about climate change last week. And it succeeded. At the Queen's official opening of parliament on 15 November, it promised legislation that will see the country's greenhouse-gas emissions slashed to 60% of 1990 levels by 2050. That is a level far beyond its carbon-reduction commitments under the Kyoto Protocol, and a longer-term programme of cuts than any other major polluting country has so far adopted.

Is the British pledge as impressive as it sounds? The legislation has not yet been officially introduced into parliament — that will happen in the coming months when the government produces a draft bill. So, although the announcement is a powerful statement of intent, the details remain vague. "How the target is defined and set, and how progress is measured and reported, are fundamental issues that are still being considered," said a statement from the Department of Environment, Food and Rural Affairs, which is charged with drafting the legislation.

Opposition politicians and environmentalists have also pointed out that although the far-reaching target is laudable, the plan isn't so tough in the short term. Critics have called for the plan to involve fixed annual cuts in emissions, rather than the five-yearly targets suggested by the government. There is optimism, however, that the target is achievable — if improvements in energy efficiency are combined with the development of low-carbon technologies.

Unfortunately the same cannot be said of sorting out what should happen in the next phase of the Kyoto agreement. International talks in Nairobi, Kenya, last week were meant to address how to incorporate large developing economies such as China, India and Brazil into the treaty after 2012. On Friday, the summit's final day, the Kyoto nations decided to postpone the negotiation until 2008.

Michael Hopkin

in the Italian research system — such as recruitment. Rules alone won't stop dishonest hiring policies, says Modica, but universities that allow bad decisions to be made will now be stung by the evaluation system.

The second decree radically changes selection procedures for presidents of research organizations. These were previously direct government appointments — former prime minister Silvio Berlusconi's centre-right administration appointed several research heads considered incompetent by much of the scientific community (see *Nature*

440, 264–265; 2006). Now independent committees will prepare shortlists of three candidates from which the research minister must pick.

The new rules are already being put into practice for the Italian space agency, whose previous president, Sergio Vetrella, has resigned under pressure from the government. The rules are also expected to make it easier for the government to oust Fabio Pistella, president of the National Research Council, which runs more than 100 research institutes around Italy. Pistella has evaded attempts to

transfer him to another post.

Scientists have welcomed the reforms. They will help a lot, even if funding levels for 2007 are modest, says physicist Giorgio Parisi of La Sapienza University in Rome. But after five successive years of cuts, he says it will be "disappointing" not to maintain at least 2006 levels.

The budget bill is now being considered by the Senate, on which Levi-Montalcini sits. She is reconciled with a relieved government whose coalition's majority in parliament is so narrow that every vote counts. ■ Alison Abbott



NORTH KOREA'S NUCLEAR FACTORY HITS SNAGS

US scientist reports setbacks at plutonium plant.

www.nature.com/news

Aboriginal remains head for home

The origins of the small collection of bones that will soon travel from London's Natural History Museum (NHM) to Tasmania say much about why the journey has taken so long to come about — and is still so controversial. One skull comes from an unnamed Tasmanian aboriginal woman, shot by a white settler and later decapitated. The curatorial description for another set of remains notes coolly that: "There has been white settlement of Tasmania since 1802. The remnants of the blacks being removed from the island in 1831."

Just a dozen or so Tasmanian aboriginal women survived the brutal colonization by British settlers, and only a handful of museums retain specimens of the now dead race. For scientists, those remains are a crucial illustration on a page in human history. But for modern Australians descended from the aborigines, they are stolen property. On 17 November the NHM trustees announced that, in this case at least, indigenous claims trump scientific value.

The decision is a landmark in a process that has seen human remains repatriated to an increasing number of indigenous groups over the past 20 years. In the United States, a change in the law has allowed native Americans to request the return of cultural and biological artefacts since 1990. But the move by the NHM could mark a change in the process, as few remains have so far been repatriated to groups outside the country in which they are held. The museum has also unsettled scientists by assessing indigenous ownership using much looser criteria than other institutions have used.

"You have to ask questions about the will of the Natural History Museum trustees to defend science," says Robert Foley, who studies human evolution at the University of Cambridge, UK.

The NHM's decision is the first to be made since Britain passed legislation in 2004 that permits repatriation claims. Descendants of the aborigines, represented by the Tasmanian Aboriginal Centre, submitted a case for the return of the remains of 17 Tasmanians to an independent panel of experts asked by the museum to weigh the ethical and scientific sides of the argument. The four-member panel recommended repatriation; the museum's trustees agreed and announced the move immediately.



Tasmania was cut off from Australia some 12,000 years ago, giving its aborigines a unique culture.

The loss of the specimens is a blow to museum scientists. Rising sea levels cut Tasmania off from Australia some 12,000 years ago, isolating people there from the innovations, such as stone tools, that changed life on the rest of the continent. NHM researchers say that the Tasmanian remains are a crucial record of the region and a vital component in records of human variability. Very few Tasmanian remains exist in other museums, and those in Australia are often not accessible to scientists.

The NHM will retain the remains for three months to sample DNA and run computerized tomography scans. Storing such scans reduces

the loss caused by the repatriation, but researchers point out that as the remains are likely to be cremated, experimental techniques that arise in the future cannot be applied. New analytical techniques were behind important results on Neanderthal

DNA published just last week, they point out (see *Nature* 444, 254; 2006). "Who knows what kind of questions we could ask," says Daniel Lieberman, an anthropologist at Harvard University.

Others worry that the NHM's decision puts at risk other foreign remains in its collection, which make up just under half of its 20,000 human

specimens. The museum is already talking to the Australian government about the return of some 450 other aboriginal remains, and has also been approached by indigenous groups from New Zealand and the United States.

Museums in Europe and the United States typically require such groups to prove a strong connection to the community from which the bones were removed. Yet the Tasmanian claim is based on a relatively loose connection, as the small number of aboriginal ancestors became part of other communities. Such a claim would not satisfy US law, say experts familiar with the process. It also seems to contradict UK guidelines, which call for descendancy to be proved. "I find it bizarre," says one researcher familiar with the guidelines, who asked not to be named. "There is no evidence that [the Tasmanian Aboriginal Centre] is the appropriate body" to receive these remains.

But Richard Lane, the NHM's director of science, says the museum made the decision after considering how the descendants of the Tasmanian aborigines view their link with the remains. He says that aboriginal groups place a great deal of importance on cultural links with their ancestors. Western thinking emphasizes genealogical descendancy above such links, but Lane says it would be unethical to place European views above those of the aborigines. He adds that the advisory panel's guidelines, drawn up after considering the national guidelines, reflect this approach.

Jim Giles



"You have to ask questions about the will of the Natural History Museum trustees to defend science."

Y. ARTHUR-BERTRAND/CORBIS

K. SAYER/CORBIS

ZOO NEWS

Penguins in boots

Penguins at the International Antarctic Centre in Christchurch, New Zealand, were getting calluses and infections on their feet from spending too much time waddling on the land. So staff have begun a vigorous foot-treatment regime, including giving the birds rubber-soled shoes.



Panda poo

A Thai zoo has been turning bamboo pulp pooped out by its two pandas into paper. According to project manager Prasertsak Buntrakoonpoontawee, the zoo has been earning about US\$8,200 a year from selling fans, cards and bookmarks made from the excrement paper — all with panda faces on of course.

NUMBER CRUNCH

String theory according to *Esquire* magazine:

20 years have passed since string theory became dominant.

10⁵⁰⁰ is the number of potential string-theory solutions.

0 is the number of testable solutions.

ON THE RECORD

“We don't want to jeopardize the iconic nature of the French fry.”

McDonald's chief executive Jim Skinner explains why the fast-food chain isn't ready to switch to a healthier oil for its signature fries.

“I feel guilty about the huge hole in the ozone layer my haircuts created. It's my responsibility to right the wrongs of the Eighties.”

Rocker Jon Bon Jovi (below) prepares to save the planet.



Sources: National Geographic, Reuters, AP, Radio Times

Universities urged to do more for poor nations

This month sees the fifth anniversary of the Doha declaration, an international agreement signed by the world's trade ministers that was aimed at broadening poor countries' access to medicines. But in the intervening five years, the main provisions in the declaration have not been used and access to treatments remains dire. So activists are changing tactics. Instead of leaning on governments and corporations, as they have in the past, they are now pressuring universities to guarantee access to drugs and medical products invented on campus.

This idea — labelled socially responsible technology transfer — has already notched up some key victories. For instance, in 2001, after student protests, external pressure and heavy media coverage, Yale University and Bristol-Myers Squibb agreed to an unprecedented arrangement. The company said it would allow companies in developing nations to produce a generic version of a key AIDS drug, d4T (stavudine), which had been invented at Yale and licensed to the company.

And last year, the University of California, Berkeley, issued a royalty-free licence on a process invented by one of its scientists, Jay Keasling, in which engineered yeast churn out the malaria drug artemisinin (see *Nature* 440, 852–853; 2006). The deal was supported in part by grants from the Bill & Melinda Gates Foundation, and was one of 15 that has been handled

by a technology-transfer programme at Berkeley dedicated to socially responsible licensing.

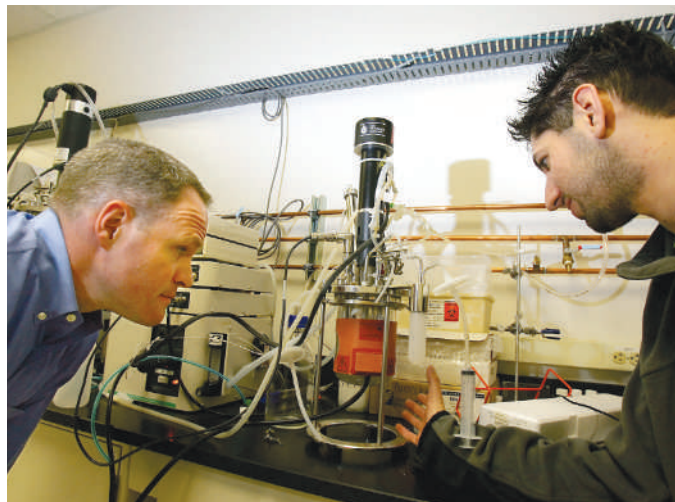
The concept of socially responsible technology transfer has also found a potentially influential political backer in US Senator Patrick Leahy (Democrat, Vermont), who is expected to head the powerful Committee on the Judiciary in the next Senate. In September, Leahy introduced a bill that would require federally funded research institutions to ensure that their drugs and novel medical devices are supplied cheaply to the developing world.

But the movement now faces several difficult tests, and the next year could see the answer to a key question: can socially responsible technology transfer gain permanent traction?

Its proponents say yes — and they are backed by a host of prominent supporters. On 14 November, a group of students called Universities Allied for Essential Medicines (UAEM) released a manifesto that it calls the Philadelphia Consensus Statement. This has been signed by some 400 scientists, lawyers, public-health specialists and students. The list boasts four Nobel laureates, as well as Paul Farmer, founder of the medical justice group Partners in Health; and the Justice Edwin Cameron, the first high-ranking South African leader to disclose that he had HIV.

The consensus statement calls on universities to promote equal access to the fruits of their

“Universities have to show that their mission statements are not just lip-service.”



Success story: the technique for generating an antimalaria drug from yeast, developed by Jay Keasling (left), has been given a royalty-free licence.

M. J. SANCHEZ/AP

FOTOS INTERNATIONAL/REX FEATURES



CALIFORNIA CAUGHT OFF GUARD BY TSUNAMI
Find out why warnings were called off, despite accurate predictions.
www.nature.com/news

SNAPSHOT

A night out in the park

These startled tigers were captured on film by a camera trap in Nagarhole National Park in India, as part of a nine-year monitoring study by biologist Ullas Karanth and his colleagues.

This month, the researchers report the astounding result: an average of one in four tigers living in the park dies or leaves each year — and yet, the tigers there are thriving. The study was conducted and analysed by Karanth's team from the Wildlife Conservation Society in India and by scientists from the US Geological Survey (K. U. Karanth *et al. Ecology* 87, 2925–2937; 2006).

How can the tigers survive despite such high annual losses? Karanth says the results bolster his hypothesis that tigers reproduce fast enough each year to make up for deaths due to poaching outside the park — as long as the big cats have enough prey to eat and are protected inside park boundaries.

Other conservationists want governments to place more emphasis on curbing the trade in tiger parts and skins, which are in high demand in



Tibet and China. But to Karanth, trade controls aren't as crucial as efforts to secure healthy habitats for the tiger. So measures such as better park patrols and voluntary resettlement of people living in parks are the best way forward, he says.

Erika Check

research, to step up work on diseases that disproportionately affect the poor, and to measure research success according to its impact on human welfare. The group has delivered its statement to an influential group of technology-transfer officers from prominent US universities. And last week, it submitted the statement to a World Health Organization working group that is mapping a more sustainable strategy to develop drugs for neglected diseases.

"Universities have this lofty language in their mission statements about creating and disseminating knowledge for the public good," says Dave Chokshi, a medical student at the University of Pennsylvania in Philadelphia and a member of UAEM. "If they want that to mean anything, they have to take up these types of issues and show it's not just lip-service."

But technology-transfer officers at universities say that it's not always easy to do the right thing. Some argue that the socially responsible technology-transfer movement doesn't understand the pressures involved in marketing university inventions to potential investors.

In a deal announced this June, for instance, Yale licensed a new AIDS drug, Ed4T, to a Japanese company called Oncolys BioPharma.

Ed4T is chemically very similar to d4T. But UAEM has criticized the deal, alleging that the licence doesn't do enough to secure access to the drug for the poor. The students claim that Yale seems to have reversed the course it set with the 2001 d4T agreement.

The Clinton Foundation HIV/AIDS Initiative, based in New York, has also questioned Yale about the terms of the Ed4T licence. In a response to the foundation on 26 October, Yale president, Richard Levin, outlined steps that the university has taken to promote access to Ed4T. For instance, the university does not plan to enforce its patents on the drug in some low-income countries, such as India. And it will not grant licences for companies that want to sell the drug in these nations, adds John Soderstrom, managing director of Yale's office of cooperative research.

Soderstrom says that Yale would have liked to do more, but that its hands were tied in this instance, adding that Ed4T was "a very difficult case". "We marketed this for years before anyone showed any interest in it," Soderstrom says. In fact, he argues, Yale is lucky if it can get one company interested in an invention — and that limits the university's ability to negotiate terms

that ensure broad access to drugs.

"We were very excited because the company was willing to make a commitment to develop the technology and bring it to the marketplace, but we clearly weren't able to get every single term and condition into the licence agreement that we would have liked," Soderstrom says.

This is why Carol Mimura, head of Berkeley's socially responsible licensing programme, says the third plank of the Philadelphia Consensus Statement is so important. Universities must be open both to new types of arrangements, and to new ways of measuring success in partnership ventures, she says. Simply looking at the current metrics of deals done and dollars earned doesn't capture the full spectrum of benefits that universities reap from socially responsible technology transfer.

"If you measure success in terms of social impact or awareness and you count things such as gifts, research collaborations, global impact and boost to your reputation, it changes your orientation," Mimura says. "If you measure success only by the amount of royalties and fees you bring in, then your licensing practices will reflect that."

Erika Check

Neuroscientist resigns over recruiting battle

A bruising clash between neuroscientists at the Massachusetts Institute of Technology (MIT) ended last week when Nobel prizewinner Susumu Tonegawa quit as head of the Picower Institute for Learning and Memory in Cambridge.

Tonegawa was accused this summer of discouraging a new female recruit from accepting a position at the neighbouring McGovern Institute for Brain Research, also part of MIT, reportedly because her work would compete with that of his lab. An investigation into the incident earlier this month concluded that Tonegawa acted inappropriately, but its report provoked anger from Tonegawa's critics, who felt that he had escaped too lightly.

His departure may soothe relations between the two institutes. The university said that it was Tonegawa's decision to step down, and that he will remain an MIT professor.

Terror law tightens on animal-rights protesters

The United States looks set to expand the reach of a 1992 law aimed at protecting animal facilities from violent activists. On 13 November, the House of Representatives passed the Animal Enterprise Terrorism Act; the Senate has already approved it, and President George W. Bush is expected to sign it into law.

The bill makes it a federal crime to use force, violence or threats against companies affiliated to animal enterprises, such as accountants and insurers, and their employees. The 1992 law explicitly covered only animal facilities themselves, from research labs to circuses and zoos. Animal activists have recently begun targeting employees of affiliated companies. The

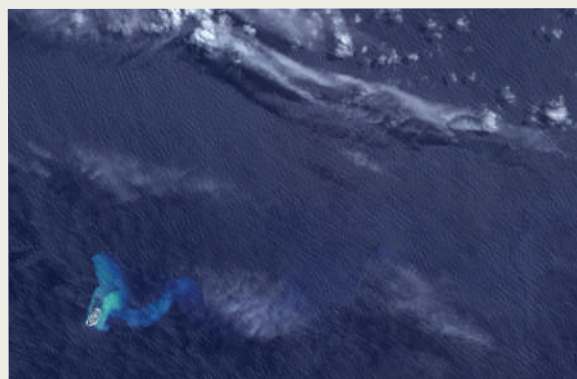
Submarine eruption bares volcanic island in Tonga

The Pacific Ocean kingdom of Tonga has a new island, but its size seems to be dwindling by the day.

In August, a passing yacht reported seeing floating chunks of pumice near the underwater Home Reef volcano. Then an entire new island emerged (lower left in the image), around 1.5 kilometres across, and complete with four peaks and a central crater. Satellite images have helped monitor the new land as it rises and recedes.

Images obtained on 12 and 14 November from NASA's Terra satellite, for instance, suggest that the island has shrunk in surface area by about a third since early October. Small volcanic lakes on the island have also disappeared.

Home Reef last erupted in 1984, when it created an island of roughly the same dimension — which also later vanished beneath the waves. Alain Bernard of the Free University in Brussels says he expects the new island to be gone within a month.



NASA

bill also adds a new criminal offence: intentionally placing a person or their family in reasonable fear of death or serious bodily injury.

The bill does not define as crimes activities such as peaceful picketing and economic boycotts. Some civil-liberties advocates have protested against the legislation.

Gene-sequencing patent challenged after 25 years

Ownership of a central invention underlying gene sequencing was called into question this month when the US Patent and Trademark Office allowed a small company to challenge the patent almost 25 years after it was filed.

The patent, covering basic DNA gel-sequencing technology, is owned by the California Institute of Technology (Caltech) in Pasadena and licensed exclusively to Applied Biosystems of Foster City, California. But Enzo Biochem, based in Farmingdale, New York, claims to have filed its patent application in 1982, months before Caltech scientists filed theirs. It says that continued amendments and rejections delayed the progress of its application, which has become public only now.

The patent office is expected to take up to two years to reach a decision.

US Senate approves deal to aid India's nuclear plans

A controversial Indo-US nuclear deal overcame a final hurdle last week when it gained the approval of the US Senate.

India's leading nuclear scientists, some

of whom had opposed the deal (see *Nature* 442, 859; 2006), say celebrations should wait until it is clear that the final legislation — expected to be signed by President George W. Bush next month — addresses their concerns. They say the bill, first mooted 16 months ago, still contains clauses that prevent India from stockpiling reactor fuel, conducting nuclear tests and reprocessing spent fuel.

India's prime minister, Manmohan Singh, says Bush has assured him that "the final outcome will be in line with Indian concerns". The deal is intended to provide India with access to nuclear technology and cheap fuel in exchange for making its nuclear programme transparent.

Supernovae expose dark matter's 9-billion-year past

Dark energy — the force thought to be causing the Universe to expand at an ever-accelerating rate — has been around for at least 9 billion years, a study shows. And the mysterious force seems to have been much the same then as it is today.

A team led by Adam Riess of Johns Hopkins University in Baltimore, Maryland, has used the Hubble Space Telescope to observe 24 old and distant exploded stars, or supernovae. Similar work on younger supernovae had revealed that the stars are farther away than would be expected if the Universe was expanding without the presence of dark energy.

The strength of the dark energy seems to have been much the same 9 billion years ago, says Riess, although the error bars on the team's results are very large. The findings will be published in *The Astrophysical Journal*.

AP/M. EVANS



Under the latest bill covering the actions of animal-rights protesters, peaceful picketing remains legal.

BUSINESS

Nuclear's core business

The job of cleaning up Britain's nuclear plants is up for auction — so who might profit from the newly privatized industry? **Andrea Chipman** reports.

Last month, the UK government announced that it will split its nuclear-site management and clean-up business. British Nuclear Group (BNG) is to divide into two separate parts, re-tendering its largest contract rather than being sold off in one piece as originally planned. This was just the latest tweak to the process of creating a marketplace for the clean-up of Britain's nuclear legacy that continues to keep the nuclear business and potential investors on their toes.

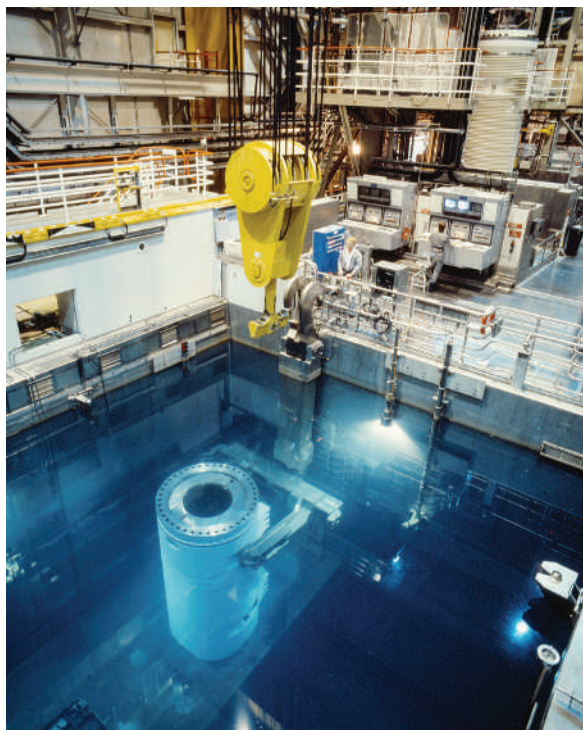
Over the next six years, the Nuclear Decommissioning Authority (NDA) — a public body created in 2005 to take strategic responsibility for 20 civil nuclear sites — will open the clean-up contracts to bidding. Other sites that will ultimately go to tender include a handful of military facilities and eight nuclear reactors owned by electricity company British Energy, which haven't yet reached the end of their lifespans.

Powering down

Sites slated for clean-up include Dounreay, a 1950s research reactor at Caithness in Scotland, Sellafield, a two-square-mile nuclear facility on the England's West Cumbria coast, and some of the first-generation Magnox power plants. It will be a huge task. The total volume of radioactive waste subject to the clean-up is forecast to total 2.3 million cubic metres, according to a 2004 inventory report by Nirex, the nuclear waste management agency.

The NDA, which is due to come up with a 'definitive' clean-up figure for the government by April 2008, currently spends £2.2 billion (US\$4.2 billion) a year on clean-up contracts. Meanwhile, the government's estimates of the cost of the clean-up have been rising by some 10% a year for the past four years and have recently topped £63 billion. But some industry insiders believe the figure is nearer £85 billion, others double that.

Sixty per cent of total clean-up costs are expected to be used up at Sellafield, which operator BNG refers to on its website as "the most challenging nuclear site management programme in the world". With huge water tanks containing radioactive waste and fuel



Deep water: Sellafield nuclear plant presents a major challenge.

reprocessing plants, Sellafield is likely to require the most technical ingenuity of contractors. "There's so much radioactive hazard there that when you work on one part, you need to be aware of everything else around it," says Paul Vallance, a spokesman for BNG.

"The opportunities are increasing for private companies, but there's more competition."

— Peter Highton

In addition to the government-owned United Kingdom Atomic Energy Agency (UKAEA), which is currently responsible for cleaning up Dounreay, a number of other companies are keen to enter bidding. US engineering giant Bechtel said earlier this month that it had formed a consortium with BWX Technologies, a nuclear facilities management firm in Lynchburg, Virginia, and UK services firm Serco Group to compete for contracts. Fellow UK services company Amec, which already subcontracts on 250 nuclear clean-up projects in Britain, has joined with UKAEA and CH2M Hill, an engineering company in Englewood, Colorado, to bid in the new market.

"The opportunities are increasing for private companies, but there's more competition,

says Peter Highton, managing director of Amec NNC, the company's nuclear division.

Less clear, say analysts, is how investors' requirements will dovetail with the government's remit to open the nuclear clean-up market to competition.

Given that the decommissioning market is one that will shrink over time, government contracts are likely to be appealing to companies only if they offer dependable profits, says Ian Jackson, an independent nuclear-industry analyst and head of UK firm Jackson Consulting. The pending break-up and sale of BNG, which holds the lion's share of nuclear clean-up contracts, is a case in point, he adds.

In a research report earlier this month, Jackson discussed the "paradox" of the government's nuclear privatization strategy: "Its largest player, BNG, is worth little to investors if it has no forward contracts, making the firm difficult to sell. But granting forward contracts that continue BNG's existing monopoly position will destroy the market diversity that the NDA is trying hard to achieve."

Investor incentives

Ultimately, Jackson adds, the component businesses of BNG are likely to be sold at a "knockdown price" that allows investors to bet on those businesses winning clean-up work without having long-term forward contracts included in the sales agreement. Another attraction for companies is to boost profits by using new clean-up technologies, says Clive White, head of Amec's UK operations.

Amec's US nuclear business, for example, has developed its own proprietary technology, Geomelt, which uses electric currents to melt radioactive and chemical wastes into a manageable glass-like material for safe storage. UKAEA has had some limited success applying marine technology by using remotely operated vehicles to detect and clean up radioactive contamination found on the sea bed and beaches near Dounreay. Other approaches, such as applying nanomaterials to the immobilization and storage of radioactive waste, are still at the early stages of development.

In theory, the long-term financial rewards for investors could be significant. The NDA's strategy paper of March 2006 says the authority will consider raising the 4.4% fee for performance that it awards contractors to as much as 15%. Under this higher calculation, Jackson notes, the Sellafield business could be worth as much as £451 million alone, with the Magnox plant clean-ups valued at more than £500 million. ■

S. ALLEN/SPL

CONSCIOUS OF CHANGE

Khotso Mokhele, formerly in charge of developing research in South Africa, talks to **Michael Cherry** about the role that science is playing in the nation's development.

It feels slightly odd meeting Khotso Mokhele in Paris, where he is attending an International Council for Science meeting. Usually we meet on African soil, but this year I am on sabbatical, and Mokhele, who is 51, stepped down in September after ten years at the helm of Africa's largest research funding agency, now known as the National Research Foundation (NRF) in South Africa. He seems quite at home, however, in the 16th *arrondissement*, domain of the city's diplomats and aristocrats. With the current resurgence of interest in Africa in the West, I wanted to hear his views on the role that science and technology are playing in the development of South Africa and the rest of the continent.

In 1992, the relatively unknown microbiologist was plucked from his department at the University of Cape Town to become the vice-president of the NRF's predecessor, the Foundation for Research Development. After serving a four-year apprenticeship he took over the reins from Reinhard Arndt. But he is the first to admit that his appointment was not a foregone conclusion. Mokhele is not a member of South Africa's ruling African National Congress (ANC) — and therefore not an obvious potential candidate to run a government agency. His tiny party, the Azanian People's Organisation (AZAPO), has strong roots in the black consciousness movement, a political philosophy born in the late 1960s that emphasizes the value of black identity.

Conflicting interests

During Mokhele's tenure, the research foundation extended its funding to social sciences and arts — it was formerly confined to the natural sciences and engineering. South African science has enjoyed mixed fortunes over the past 15 years: although spending on research and development (R&D) as a percentage of gross domestic product (GDP) has risen recently to 0.81%, it fell drastically in the nineties from a peak of 1.03% in 1991 to a low of 0.68% in 1997. But studentships and postdoctoral fellowships are worth less in real terms, making recruitment into R&D more difficult.

But what does Mokhele regard as his biggest achievement of the past decade? "We managed to retain quality as a cornerstone of the research enterprise," he replies, with very little hesitation. The South African science

community is mostly white, and Mokhele was pressed to boost the number of black scientists. "A research culture can die very quickly if you uncouple it from quality considerations. As a black South African, I would have been excused if I had addressed equity issues in a manner that resulted in the disappearance of quality."

This is undoubtedly true. Eight hundred thousand white South Africans — about 15% of the white population — have left the country over the past decade, citing inequality of opportunity as a contributing factor. Doctors, accountants, engineers, teachers, nurses — but very few researchers. This is something that could be regarded as a tribute to Mokhele's efforts: for example, the NRF awards grants to the best research projects, as well as providing incentives for black researchers.

"The crime is that 12 years after the advent of democracy, black education has made no strides, so we are still reliant on the white community for these skills," says Mokhele. "Whites may complain about affirmative action, but in reality they experience very little competition for jobs because of this lack of progress."

The seeds of his passion for education and valuing black identity were sown during his upbringing. Mokhele grew up in the township of Phahameng in Bloemfontein, capital of the Free State province. His mother ran a *shebeen* (an unlicensed pub) in their house from Thursday to Sunday. Her vocation had an unexpected

benefit: as a young boy, Mokhele became an expert brewer, using a paraffin tin in the backyard, ultimately inspiring his interest in microbiology. "I had the knack: when my brothers tried to do it, they got the temperature wrong."

Race relations

At the age of 16, Mokhele was sent to Moroka High, a mission school in Thaba Nchu in the eastern part of the province. His chemistry teacher suggested he apply to the University of Fort Hare in Alice, Eastern Cape province — alma mater of Nelson Mandela, Mangosuthu Buthelezi and Robert Mugabe — to study agriculture. "As a township boy, agriculture was the last subject I wanted to study," he laughs, "but I did it as a last option." By the end of his first year, he was hooked — not just on food science, but also on politics, as he had come under the influence of the black consciousness leader Steve Biko.

Biko lived in nearby King William's Town, and, although banned from Fort Hare by the university's reactionary administration, he frequently visited the adjacent Federal Theological Seminary, where in Mokhele's words: "We listened to him think aloud. There were also a lot of parties, and Biko would leave the dancefloor to write down his thoughts when he had a moment of inspiration."

At the end of his third year, Mokhele was expelled from Fort Hare for political activity.



Identity politics: Steve Biko's thoughts on black consciousness are at the heart of science policy in South Africa.



"The crime is that 12 years after the advent of democracy, we're still reliant on the white community for skills." — Khotso Mokhele

was not obvious from the outset, and we failed there." South Africa is currently providing antiretroviral treatment to only 13% of its inhabitants requiring it, compared with more than 50% in neighbouring Botswana and Namibia, for example.

As for the future, Mokhele believes that South Africa should play to its strengths.

The country is known, for example, for its contributions to oceanography, astronomy, geology and ecology. "There is little point in the country making tiny contributions to the same fields of science that are currently fashionable in the West. We need to do two things: concentrate on areas where we have a natural advantage; and interpret what is going on elsewhere and adapt it to our specific needs," he says.

Continental shift

I turn to Africa as a continent — why has it fared so badly in comparison with other post-colonial societies? "Apart from Botswana, non-military democracy in Africa is very new. Africa is also the only continent where most children study in a language that they don't use when they play in the streets."

And what about the mooted concept of the African Science and Innovation Facility — could an Africa-wide research agency work, in reality? He's not sure: "There has been such a reluctance on the continent to establish national agencies for research funding, and it is the strong nations that stand to benefit from globalization. The only way that such an agency could work would be if beneficiary countries were obliged to spend a certain percentage of their GDP on research in order to qualify. If it's based just on hand-outs from the West, it is destined to fail."

Mokhele's career to date has benefited from his natural charm, political savvy and the odd measure of good luck. Robert Kriger, who worked closely with Mokhele at the NRF over this period, says that Mokhele's great strength "is his ability to get to the core of research issues — which reflects both his experience within the international science community, and his knowledge of the workings of its associated bureaucracies".

When asked about his own future, Mokhele describes himself as "being in retirement until such time as I decide what to do next". But he is not tempted to follow so many of his compatriots out of Africa, a position to which he remains unequivocally committed. ■

Michael Cherry is *Nature's* contributing correspondent in South Africa.

But like many adherents of black consciousness, Mokhele has a strong sense of his own destiny. Biko's philosophy, in a nutshell, was that black people needed to develop their own sense of identity, rather than take their cue from white, Indian and mixed-race intellectuals who dominated the democratic movement at the time. "I wasn't going to let my expulsion determine my fate," Mokhele says. This self-assuredness is an attribute that has made him able to work effectively with both black and white communities.

Fort Hare agreed to readmit him to complete the degree and he left South Africa in 1979 for the University of California, Davis, where he completed master's and doctoral degrees. But in 1987, he returned to Fort Hare to a temporary lectureship. "It was a tough decision," says Mokhele, "as I felt very comfortable in the Afro-American community in the States."

Looking back over his tenure at the NRF, Mokhele says there isn't really anything he would have done differently, even the decisions that didn't turn out as well as he had hoped. He served, for example, as a facilitator for President Mbeki's controversial advisory panel set up in 2000 to investigate the link between HIV and AIDS. Many panel members were 'dissidents', set on proving that there was no link. No one envied Mokhele's task of trying to achieve even a measure of consensus between these dissidents and an enraged and alarmed group of orthodox scientists.

The planned experiments on which the panel had supposedly agreed were never concluded and Mokhele's critics felt he was left with egg on his face. "I had no alternative but to accept that job," he counters. "It could have played itself out in a way that would have saved us from the situation we are in now. It didn't, but this outcome

A MEASURE OF HAPPINESS

Philosophers since Aristotle have puzzled over the meaning of happiness. **Tony Reichhardt** asks what scientists, psychologists and economists can bring to the topic. Are we any closer to being able to quantify joy?

Feeling happy? As you read this, are you taking a well-deserved break from your work, confident that your latest experiment is going to produce the results you want? Or is reading *Nature* a guilty pleasure, snatched in the face of other pressures — your students, your department head, writing the next grant?

Sociological surveys usually try to measure well-being by asking people to assess their current level of happiness. Such questions are a regular feature of surveys of the public, such as the 2005 Pew Research Center survey, which asked 3,014 adult Americans: “How happy are you these days in your life?” Half reported being “pretty happy”, 34% “very happy” and 15% “not too happy”.

But asking people if they are happy raises more questions than it answers, not least of which is how to define happiness. Is it a single emotion or a personality trait? A physical state, with characteristic brain-wave patterns and biomarkers? Is it simply the absence of unhappiness, or something else?

Psychologists, economists and other well-being researchers don't have definitive answers, but they're beginning to approach the subject in a more rigorous way. In the process, they hope to learn more about the link between health and happiness, and to contribute to debates over eternal questions, such as: who's happier, the Americans or the French?

Daniel Kahneman, a Princeton University psychologist who won the 2002 Nobel Prize for Economics for applying psychology to decision-making in the face of uncertainty, wants to develop surveys that ask more sophisticated questions. His work investigates how a person's sense of overall life satisfaction diverges from their everyday ups and downs. Early results suggest that the two do not necessarily correlate.

Kahneman and his colleagues gave two questionnaires to women in Columbus, Ohio, and Rennes, France. The first assessed overall life satisfaction. The second asked for a diary of a single day, broken into discrete episodes (had lunch with a friend, did chores), and to rate specific emotions they'd felt during each episode, on a scale from 0 to 6.

This Day Reconstruction Method (DRM)¹, developed by Kahneman, Princeton economist



Happy hour: how do feelings from moment to moment relate to overall satisfaction with one's life?

Alan Krueger and others in 1994, is easier to administer, but yields results similar to the more intrusive Experience Sampling Method, where subjects are interrupted by a beeper at various points during the day and asked to report their feelings.

Back to normal

Kahneman found some differences between French and American women — for example, Americans spent more time on childcare and enjoyed it less — but in general, similar things made them happy (time with friends and family) and unhappy (commuting and work). And in both cases, overall satisfaction was not strongly correlated with experienced happiness. Rich and married women reported more satisfaction than poor and single women, for example, but weren't happier day-to-day.

Although preliminary, these results provide a new way to investigate the widely held idea that our level of happiness is fixed. When social scientists compared survey reports of happiness with other factors, ranging from income to

pet ownership, they concluded that people with more money report being happier, but only up to a point. Beyond a certain level (for Americans in 2004, an annual household income of \$50,000 to \$90,000), more money doesn't bring more happiness. And in several countries where median income has risen — including America, China and Japan — happiness levels remain stubbornly constant.

These and other data led well-being researchers to conclude that life circumstances don't have much affect on long-term happiness. Surveys show that happiness increases after marriage, but only temporarily. An oft-cited 1978 study found that, a year after their life-changing event, both lottery winners and paralysis victims had reverted to close to their former level of happiness². This contributed to the notion of a 'hedonic setpoint' to which people return no matter what life throws their way. And based on studies showing similar levels of reported happiness in twins, the setpoint appeared to be genetically determined.

Pioneers of 'positive psychology' such as

J. BELL/SPL

Martin Seligman of the University of Pennsylvania in Philadelphia developed research-based practices for being happier at the margins — by consciously expressing gratitude, for example. But most experts agreed that we were stuck at our natural setpoint.

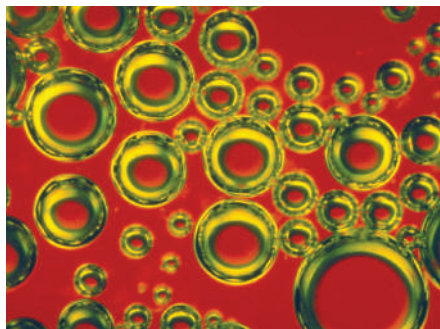
Lately, however, psychologists such as Kahneman and Ed Diener, from the University of Illinois at Urbana-Champaign, have backed away from too rigid an interpretation of this 'hedonic treadmill'. Diener argues that although genetically determined personality factors may predispose people to a certain level of well-being, different kinds of happiness change over time³. For example, older people report declines in both positive and negative moods. In Diener's view, happiness is not a unitary concept with a single set point.

In the blood

Kahneman thinks a greater focus on measuring experienced happiness could change our view of the hedonic setpoint. The observation that married people revert to their previous levels of satisfaction within a few years of the wedding may be explained by marriage adding some sources of daily happiness (time with spouses and children) while decreasing others (time with friends). Future surveys may need to distinguish between daily and long-term happiness more carefully.

In Kahneman's quest for a measure of happiness that "economists can respect", he and Krueger came up with a U-index based on DRM reports of positive and negative episodes. The U-index is the fraction of time a person is unhappy, and relates more to experienced happiness than to life satisfaction⁴.

Krueger has been working with the Gallup Organization to come up with a five-minute version of the DRM, making it suitable for large surveys. Measures such as the U-index



Spit spot: hormones in saliva might reveal a link between physical health and mental well-being.

averaged for the general population, he says, could be incorporated into National Well-being Accounts — already used in Canada and Australia — that try to measure quality of life. Krueger, among other economists, argues that well-being is a better social goal than economic growth. But first they need to agree on a simple measure of well-being.

Well-being researchers are also going beyond questionnaires to borrow tools from medicine, neuroscience and genetics. Kahneman plans to supplement his DRM data with measures of cortisol, a hormone associated with stress, to see how it varies with mood. According to psychologist Carol Ryff, director of the University of Wisconsin's Institute on Aging in Madison, there's currently a boom in well-being studies measuring cortisol.

Ryff is responsible for perhaps the most ambitious well-being research project ever conducted, the Midlife in the United States, or MIDUS II, study. With \$28 million in funding from the National Institute on Aging, MIDUS II is five years into a six-year study to assess adult Americans' health and well-being. As well as the psychological and personality surveys collected in MIDUS I, conducted in the

mid-1990s, researchers will gather data on several biomarkers.

By measuring factors such as hormone levels (cortisol or epinephrine), blood pressure or immune-system functioning, scientists hope to learn more about the physiological underpinnings of psychological well-being and distress. Studies showing that people with positive emotions catch fewer colds⁵ suggest a link between happiness and physical health.

Cut the fluff

Some biomarkers, including measures of brain activity and genetic factors, have been correlated with well-being in previous studies. In the MIDUS II study, twins and siblings will yield information on the role of genetics in well-being. And the project plans a smaller version of MIDUS II in Japan, to gather data from another culture. The researchers hope to tease out the role of some of these biomarkers as they process their results.

Ryff hopes to find out whether well-being and ill-being (depression and so on) have distinct biological correlates, or whether they are at opposite ends of the same psychological spectrum. One of her previous studies on a group of 135 older women assessed on biomarkers such as cortisol and waist-hip ratio, suggested that the biological correlates of well-being and ill-being are largely distinctive⁶. That's one issue MIDUS II will investigate more fully.

In her own work Ryff likes to distinguish between hedonic well-being (moods and feelings) and eudaimonic well-being, which is more concerned with factors such as having purpose in life, continued personal growth and development, and good relationships with others. In fact, Ryff rarely uses the term 'happiness'. Perhaps that's because the more scientists learn, the less precise the term has become.

That's roughly where the science of happiness stands right now — still wrestling with its own terminology. Ryff has little time for the fluffier aspects of positive psychology, which she dismisses as "a lot of PR". But one thing she and other well-being researchers can agree on is the nature of the question, "We're going after it in a serious way," she says. "In the final analysis, it's an empirical question."

Tony Reichhardt is a science writer based in Fredericksburg, Virginia.

1. Kahneman, D. *et al. Science* **306**, 1776–1780 (2004).
2. Brickman, P., Coates, D. & Janoff-Bulman, R. *J. Personal. Social Psychol.* **36**, 917–927 (1978).
3. Diener, E., Lucas, R. E. & Scollon, C. N. *Am. Psychol.* **61**, 305–314 (2006).
4. Kahneman, D. & Krueger, A. *J. Econ. Perspect.* **20**, 3–24 (2006).
5. Cohen, S. *et al. Psychosomatic Med.* **65**, 652–657 (2003).
6. Ryff, C. D. *et al. Psychother. Psychosomatics* **75**, 85–95 (2006).

See Editorial, page 401.

D. CHENOWETH/III-COLUMBUS/CORBIS SYGMA



Passing mood: lottery winners and accident victims both seem to revert to their prior levels of happiness.

SPYING ON NATURE

The biggest project in the history of ecology is nearing its dawn. Can its organizers pull off the seemingly impossible and unite a disparate field behind its vision to observe the ecosystems of the United States? **Michael Hopkin** reports.

They call it the Hubble telescope of ecology. Like astronomy's best-known space telescope, the National Ecological Observatory Network (NEON) is planned to be big, expensive and able to see most of its universe. NEON, the biggest undertaking in the field's history, would blanket the United States with a network of observation stations — providing the most detailed picture ever of the state of its ecosystems.

But like the Hubble telescope, NEON and its organizers have faced a long and bumpy ride. They have fought off long-running charges that the project is too expensive, too inflexible and too poorly designed. Now they are hoping for the same sort of miraculous turnaround that the Hubble has experienced several times — to launch their project on time and with the full support of the ecology community.

Funding for NEON's construction remains at the discretion of Congress — the National Science Foundation (NSF) has requested \$12 million as part of its 2007 budget for building major research installations, rising to a total spend of \$100 million by 2011. Other costs, such as operating and maintaining the project, are expected to more than double this figure.

Grand challenge

Meanwhile, the project's members have just finished presenting the latest version of the project's blueprint to a board of NSF officials. Should their plans meet with the approval of the board — who have experience in evaluating huge projects such as telescopes and particle accelerators — the team will proceed in making the preliminary design a reality.

According to the latest blueprint, NEON will involve setting up permanent measuring stations at locations in each of 20 'domains', which split the entire United States into zones representing every major climate and ecosystem type. Each location will boast an array of instruments, including a 10-metre tower to measure carbon dioxide flux, detect nitrogen oxides and record leaf wetness. The goal — the 'grand challenges' of ecology — is to assess how ecosystems respond to natural and human-induced changes in climate, land use and the presence of invasive species.

It won't be easy. As the most expensive

project in ecology's history, NEON has struggled for funding for years. Part of the problem is that it's difficult to put an overall price-tag on what is still a work in progress. "It's always in the hundred-million-dollar range," says Dan Johnson, communications director for NEON. "But it's frustrating because it's so hard to cost out." To make matters worse, early designs were poorly received by the community, and project organizers were accused of taking plans behind closed doors. Even now, some ecologists worry the project's vision still has not been clearly enough stated, despite having been on the drawing board for more than a decade.

The project was first dreamed up in the mid-1990s by Jim MacMahon, now board chairman, and Bruce Hayden, an environmental scientist at the University of Virginia in Charlottesville. But it wasn't until 2003 that the NSF began to

include NEON in its formal plans. Originally, NEON organizers pushed to have the system as standardized as possible, with identical hardware set up at some 60 locations. Each of the 20 domains was proposed to have three permanent research stations — one in a rural area, one in an urban one and a third in a transitional zone. But the design allowed very little scope for portable measuring equipment or flexibility to address local ecological issues on an ad hoc basis — such as the risk of wildfire in a region.

Chris Field, an ecologist at Stanford University in California, helped lead a team in writing NEON's latest blueprint to address such concerns. "If there was a consistent criticism, it was that earlier designs didn't present an inviting enough edifice to ecologists," he says.

Field also takes up the Hubble telescope analogy to explain why NEON needs local



flexibility. “With a major telescope,” he says, “you can improve flexibility by ensuring that it can point at different regions of the sky, or have different instruments built on.” Similarly, ecologists want to be able to adapt their machinery to hit local hot spots of interest — such as a regional drought, or a flood such as that in New Orleans following Hurricane Katrina — with mobile measuring devices to improvise their studies.

Progress report

The latest blueprint seems to be soothing a lot of nerves and egos. In an open letter accompanying the document, MacMahon said that it is “gratifying to put to rest the gloom and doom that I have heard from some in the community”. And until recently the worries and doubts were still very much at the surface. As recently as August, ecologists packed a conference room for a somewhat irritable ‘village meeting’ that updated NEON’s progress. As one attendee put it: “Is this still going to attract money? Please tell me so I know how badly I’ll sleep tonight!” But since the release of the latest blueprint, researchers in the main seem to be more determined to make the most of the new opportunities.

The new blueprint involves just a single ‘core site’ in each domain, located within a watershed area and ideally featuring a perennial river or

stream. Each site will be equipped with a tower fully kitted out to measure gas fluxes, water vapour, solar irradiation, ozone concentrations, levels of pollen and bacteria, as well as air and soil temperature, wind speed and direction, precipitation and atmospheric pressure. Each site, which will take measurements from a few tens of square kilometres, will also feature terrestrial and aquatic sensors scattered over a wider area to measure temperatures and nutrient levels.

Apart from the core sites, NEON also includes a range of mobile instruments to measure specific local soil or aquatic processes; access to Earth-observation satellites and aircraft to monitor vegetation from above; and a ‘land-use package’ that will allow researchers to address issues such as the encroachment of new roads and agriculture into pristine land.

It’s certainly a lot for ecologists to take on board. “I think the scientific community is generally supportive and encouraging,” says ecologist David Briske of Texas A&M University in College Station. “But the flip-side is that it’s unprecedented and there’s no model.”

Hence the feelings of frustration, among some in the community, that the whole

endeavour may be getting too big for itself. Elizabeth Blood, one of NEON’s organizers, tries to calm some of those fears. “We have never tried to do this before, to enrich biological science in this way,” she says. “We are creating new ways of doing science — ways we can get only a glimpse of now.”

In its new incarnation, NEON has the flexibility to address questions not yet dreamed of, says James Collins of the NSF, another member of the project team. “Projects can go for 10, 20, 30 years, but then after that people will be proposing new research,” he says. “People who aren’t even born yet will be developing new idiosyncracies for it.”

Many ecologists are finally allowing themselves to become

enthusiastic over NEON. Perhaps most tellingly, the leaders of the Consortium of Regional Ecological Observatories (COREO), a group set up to represent the ecologists who will use the infrastructure, now seem at peace with the NEON board. “Before, there was surprise and dismay at the inflexibility, and at the fact that there was little option for COREO members to comment. But most members are pleased with the re-review,” says COREO chair Phil Robertson of Michigan State University in East Lansing.

“We have never tried to do this before, to enrich biological science in this way.”
— Elizabeth Blood

Nature watch: a network of stations will measure the impact of human activity and invasive species on ecosystems on a huge scale.

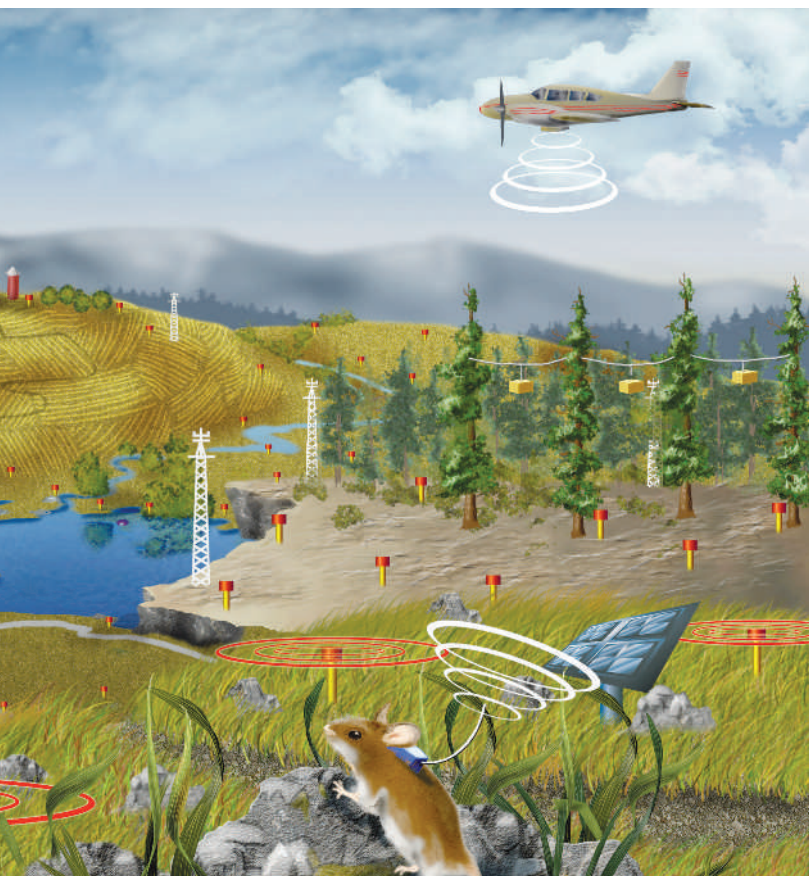
Lingering concerns

Most concerns revolve around the ambition of the project’s timeline: the system is planned to be up and running by 2013. NEON’s public-relations team insists that the project is on schedule and, although Collins admits that this is a big task, he’s confident it can be done. “It’s ambitious, but the community has been involved in the research and development for years,” he says. “The ideas have been proofed in concept — now we’re at the point of knitting them together.”

Although the grand-challenge questions are a useful guide, it is ultimately up to ecologists themselves to design the research that will be done. “The nuts and bolts are not in place yet,” says Collins. “But these ecologists are the top investigators in their fields, so we want them to write the proposals.”

So the debate about what questions the system will answer looks set to remain in the air for a while to come. The NSF is currently in the process of a ‘request for information’ — consulting ecologists on the specific kinds of research they think they can perform with the new equipment. That process won’t be complete for a few months. “Until the request for information comes out, all bets are off,” says Robertson. America’s ecologists aren’t quite ready to breathe a sigh of relief yet.

Michael Hopkin writes for *Nature* from London.



The science community must unite over Iraq

SIR — Members of the international scientific community are deeply disturbed by the systematic murder of several hundred academics in Iraq, the recent mass kidnapping at the Ministry of Higher Education's Research Directorate (see "Gunmen seize academics at Baghdad ministry" *Nature* **444**, 252–253; 2006) and the consequent threat to close all the universities in Baghdad.

Recent assassinations have included that of Isam Kadhem F. al-Rawi, a geology professor at the University of Baghdad. Al-Rawi refused to join the estimated 2,000 Iraqi scientists thought to have fled the country (see *Nature* **441**, 1036–1037; 2006). Instead, he remained and was a candidate for minister of higher education and scientific research. He monitored and publicized the plight of academics in Iraq and had compiled and maintained a list of his colleagues who were victims of hit squads — a list that al-Rawi now joins.

Local and international authorities must ensure that scientists and academics in Iraq are protected. In July, ICSU, the International Council for Science, issued a statement condemning the violence in Iraq (www.icsu.org/9_latestnews/latest_19.html). ICSU's new Committee on Freedom and Responsibility in the conduct of Science, of which I am chairman, now urges international organizations to join ICSU in searching for ways to bring this aggression against Iraqi scientists to an end, and to express solidarity with them.

Science is a global endeavour. When its fundamental values and principles are undermined and scientists are discriminated against or maltreated, it is not only science, but society as a whole, that suffers.

Bengt Gustafsson

Department of Astronomy and Space Physics,
Uppsala University, Box 515, SE-751 20,
Uppsala, Sweden

German societies want to keep an international voice

SIR — In the News in Brief story "German science academy stumbles at birth" (*Nature* **444**, 14; 2006), it is reported that I wrote a "last-minute protest letter" opposing plans for the establishment of a national science academy. However, the Max Planck Society, together with all major German research organizations, supports the current plans for setting up such an academy.

The representative role of the proposed academy on an international level had been extensively discussed. It had been

agreed that, for any specific case, German science should be represented by the organization with either the highest competence for the issue or the best organizational fit to the international body in question. This implies that some representational activities will stay with a funding council or a research organization.

It therefore came as a surprise to the research organizations that the states commission had been advised to constitute the academy as a body with a general power to represent German science internationally.

As a consequence, the German Research Organizations wrote a clarifying letter to the states commission, which — as president of the Max Planck Society — I signed also on behalf of the heads of Germany's research-funding agency the DFG, the Fraunhofer Society, the Leibniz Association, the Helmholtz Association of German research centres and the German Rectors' Conference.

Peter Gruss

Max Planck Society, Hofgartenstraße 8,
80539 München, Germany

Imagine projects with a strong emotional appeal

SIR — Your Editorial "To build bridges, or to burn them" (*Nature* **443**, 481; 2006) concludes that a large group of environmentalists "might be more likely to change their minds about science if its practitioners would desist from sneering at emotional argument and demonstrate that science is a window through which we can see our world more clearly".

I would go further, and argue that scientists would do well to use emotion, based on sound science, in that demonstration.

Our Imagine programme (www.foundation-imagine.org) involves scientists and high-school students working together on tangible projects for application in developing countries. Scientists submit their ideas, from which small groups of students aged 16 to 18 each choose one. The student groups then draft a business plan, carry out experiments, obtain local information and outline budgets, while considering social issues.

The winning team sees its plan carried out: in 2004 the production of biodiesel from algae in Mozambique; in 2005 extraction of avocado oil in Kenya; and in 2006, redevelopment of a plantation in Suriname to produce colouring compounds.

Imagine's emotive collaboration of schoolchildren and scientists working on applications in developing countries

appeals to the media, and hence to a very broad audience that might not otherwise be reached.

Patricia Osseweijer

Kluyver Centre for Genomics of
Industrial Fermentation,
Delft University of Technology,
Julianalaan 67, 2628 BC Delft, The Netherlands

Religious authorities overrule scientists in Iran

SIR — Your Editorial support for the recent rise in scientific productivity in Iran is admirable ("Revival in Iran" *Nature* **442**, 719; 2006).

However, I disagree with the comments made by the country's health minister Kamran Lankarani in Correspondence ("Iran seeks nuclear power to replace reliance on oil" *Nature* **443**, 906; 2006) about how to help those scientists who struggle to perform quality research there. The minister did not comment, for example, on President Ahmadinejad's recent call for students to demand the removal of liberal and secular university lecturers (see news.bbc.co.uk/1/hi/world/middle_east/5316634.stm).

The theocracy ruling Omar Khayyam's country is no friend of free scientific inquiry. Consider this: each year, the month-long Ramadan fast is meant to end with a festival at the next new moon. But the fast cannot end until the Shi'a authorities have made an official observation of the lunar crescent to check the astronomers' calculations. This year, for some reason, the official observation could not be carried out in time, so religion's mistrust of science paralysed the country's social life for several days.

Kamran Behnia

Ecole Supérieure de Physique et de Chimie
Industrielles, 10 Rue Vauquelin,
75005 Paris, France

One small point

SIR — Will the nano hazard symbol reported in News in Brief (*Nature* **443**, 737; 2006) come with a free magnifying glass with which to read it?

Rupert C. Marshall

Institute of Biological Sciences, University of Wales,
Aberystwyth, Edward Llwyd Building,
Penglais, Aberystwyth,
Ceredigion SY23 3DA, UK

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

BOOKS & ARTS

Life in the universal porridge

What were the chances that the conditions in the Universe would be just right for life?

The Goldilocks Enigma: Why Is the Universe Just Right for Life?

by Paul Davies

Allen Lane: 2006. 360 pp. £22

To be published in the US in April as *Cosmic Jackpot* (Houghton Mifflin).

Jim Al-Khalili

The biggest question of all, according to writer Douglas Adams, is the one about the meaning of Life, the Universe and Everything. Paul Davies, in his book *The Goldilocks Enigma*, phrases it as “How come existence?”, but it is the same age-old question, and we humans have taken several routes to try to answer it. The option preferred by most scientists is a cop-out: that we are just a cosmic accident, and that the Universe just happens to be right for life to have evolved in it. They argue that trying to find deeper meaning behind this incredibly unlikely outcome is futile, and hope that some as yet undiscovered laws of physics (such as may one day be provided by string theory or another ‘theory of everything’) will show us that things simply could not have been any different.

There is a second option. A growing band of physicists has argued that our universe is not unique but rather one of many within a larger Multiverse. Given the quite stupendous odds against us being here and the degree of fine-tuning in the laws of nature necessary for life to have evolved, it is not surprising that some have argued for the reality of multiple universes.

Here is a simple example showing how this helps. If millions of people buy lottery tickets, it is not surprising that the winning numbers should come up (maybe more than once). Of course, if your ticket contains those winning numbers, you might consider yourself incredibly lucky. But to anyone who doesn't know you, it is nothing remarkable; after all, someone has to win. Now let us suppose that in the week you bought your lottery ticket, no one else bought one, and yet you have the winning numbers. Is this now more remarkable? For you, there is no real difference, because the odds on you winning are the same whether anyone else buys a ticket or not. But from the point of view of the set of winning numbers, it is truly remarkable that they match the numbers on the only lottery ticket. This is the dilemma we face when we contemplate the incredibly tiny odds on our existence. We are those lottery numbers, and we need a lot of tickets to be bought if we



The Hitchhiker's Guide to the Galaxy asked about Life, the Universe and Everything.

are to stand a reasonable chance of matching anyone's ticket.

For years, the issue had been dismissed anthropically: if the conditions in our world were not ‘just right’, we wouldn't be around to debate the issue. However, it is much more natural and logical to conclude that even though we might be unique, our universe cannot be. If the Multiverse contains an infinite number of universes, some of them must have the right conditions for life. Can it really be that simple?

The first half of *The Goldilocks Enigma* is standard fare, covering all the usual topics, from the Big Bang, inflation and high-energy particle physics to theories of everything and where the laws of physics come from. Davies has a knack of making these subtle and complex concepts seem straightforward. But it is the second half of the book that readers will want to skip to. It is here that he faces head-on the question of why our universe is just right for us, and he covers all the main arguments thoroughly and shows up their shortcomings. Eventually, he chooses a different path that does away with luck as well as the Multiverse. But as Deep Thought, the computer in Douglas Adams' *The Hitchhiker's Guide to the Galaxy*, says, you are not going to like it.

Davies' first suggestion is that the ‘biofriend-

liness’ of the Universe may be due to some as yet undiscovered ‘life principle’, built into the laws of physics from the very beginning, that has steered and constrained the Universe towards producing life. I find this idea hard to swallow and I don't think Davies dwells on it long enough to really make a convincing case.

He then invites us to consider a more interesting — I hesitate to endorse it with the term ‘appealing’ — idea originally expounded by physicist John Wheeler. It takes one of the weirdest features of quantum mechanics and pushes it to its logical conclusion: that conscious observers bring about the universe they find themselves in by the very act of observing it, thereby dragging it out of the quantum superposition of all possible paths it could have followed. Actually, I think this is related to what supporters of the Multiverse version of quantum mechanics would argue — with the difference that, for Davies, our universe is the only one.

The main options, then, are: first, that the Universe is a fluke; second, that it is one of many and happens to be, much like Goldilocks' porridge, just right for us; and third, that conscious observers bring the universe they inhabit into existence simply by observing it, although their teleological actions would have to reach back into the past, forcing the

right conditions to be selected at the Big Bang.

Just when the reader feels that Davies is losing his grip and sliding inexorably towards fantasy, he takes a well-timed reality check, reminding the reader, and himself, that in order to address the question of “How come existence?”, one must either play it safe and back away from the question, or be quite radical. Many physicists will not like this book, but I think it will cause the biggest stir since Roger Penrose wrote *The Emperor's New Mind* (Oxford

University Press, 1989). Most of the ideas are not new, but Davies is courageous, entertaining and persuasive in laying them out clearly. Many scientists might feel that the subject matter, as Davies acknowledges, should be “left to the philosophers and priests”, with scientists tackling only those questions they can hope to answer. But it's still a thoroughly good read. ■

Jim Al-Khalili is professor of theoretical nuclear physics at the University of Surrey, Guildford, Surrey GU2 7XH, UK.

A giant leap?

Dark Side of the Moon: The Magnificent Madness of the American Lunar Quest by Gerard J. DeGroot

New York University Press: 2006. 352 pp.
\$29.95

James R. Hansen

Dark Side of the Moon stands, and falls, on its cunning soundbites. With a cheeky rhetorical flourish, Gerard DeGroot, a history professor at the University of St Andrews, UK, attacks the integrity of the American Moon-landing programme of the 1960s. Thankfully, he does not go so far as to suggest that the landings never really occurred, being merely part of a US government conspiracy to fool the Soviet Union about the United States' strategic capability in the Cold War, but the book's presentation is almost as fanciful.

Serious research into the embedded agendas of the US space programme is well worth undertaking, and many of the questions raised by this book are worthwhile. To what extent did myths constructed by the administrations of John F. Kennedy and Lyndon B. Johnson, and sustained by NASA, manipulate the US drive to the Moon? What do the Apollo missions tell us about the growth of the technocratic mentality? How did nostalgia for Apollo corrupt America's subsequent space efforts? Several books have profitably explored such matters, notably Walter A. McDougall's *The Heavens and the Earth* (Basic Books, 1985). Certainly there is more to learn. DeGroot could have added fresh insight if only he had conducted substantive research instead of filling his book with baseless assertion and arm-waving polemic.

The most grievous defect lies not in what DeGroot does not understand about the US lunar programme, but rather what he could have easily ascertained, but did not, about the history of the Soviet programme. His indictment of Apollo

hinges on the implicit notion that the Americans were the only ones racing to the Moon. But as Asif A. Siddiqi's groundbreaking *Challenge to Apollo* (NASA, 2000) explained, the Soviets made a concerted effort to beat the United States to the Moon. Siddiqi's monumental 1,000-page book is one of the most important space-history books ever published. DeGroot fails to cite it and seems to be unaware of its existence.

The author is more adept at turning a phrase than erecting a solid edifice based on original source materials and exhaustive research. He states oh-so-matter-of-factly that the words uttered by Neil Armstrong when stepping off the Eagle were “canned” — a “freeze-dried slogan”, a “meaningless” statement, “prepared earlier and then taken into space along with

the Tang and the tubes of hamburgers”. With just a little research the cynical author could have discovered that Armstrong's “one small step” was not in the least bit canned; inside the Lunar Excursion Module, the commander of Apollo 11 formulated it in the hours between landing and stepping down off the ladder. Nor did Apollo 11, incidentally, carry Tang or tubes of hamburgers.

For DeGroot, the Moon landing marked a “high point” in “America's love affair with science and technology”, but also a decline in so much else about US society. Savouring the irony, he juxtaposes the capacity of Americans to build the sophisticated machinery needed to take them to the Moon with their inability any longer to build a decent car. While Ford Pintos exploded and AMC Gremlins fell apart, he says, Americans “arrived at parties to celebrate the lunar landing in Toyotas, Datsuns, Volkswagens, and Renaults” — another great soundbite, but one lacking fuel because, in 1969, imports made up a small fraction of the US car population, with Renault barely a blip on the screen.

In his preface, DeGroot relates that the Americans who walked on the Moon were his childhood heroes but, unlike the rocket scientists and spaceflight enthusiasts, he grew out of it, never succumbing to the sexual appeal of the rocket ship: “The tall, slender phallic tube sits on its pad while men who yearn for youth trade in techno-babble. The adventure appeals to most boys, some men, very few girls, and almost no women. Freud probably had a lot

to say about this sort of thing.”

DeGroot is not the first to make the phallus comparison; others have argued that the language of engineering and rockets has been particularly sexist. But he does not reference any of those analyses, either. Nor did he look for data that could have shown that his categorical statement about women and spaceflight is shaky.

Having previously published what he describes as a depressing book on the history of the atomic bomb, DeGroot claims he hoped for spiritual uplift by turning his attention to a history of the Moon landings. But forlorn was he to find, in place of his childhood heroes, “a gang of cynics, manipulators, demagogues, tyrants, and even a few criminals”. Apollo was not a high-water mark for our species, he says, but “a brilliant deception, a glorious swindle”.

“Hubris took America to the Moon” is DeGroot's thesis, but all he really proves is that hubris sometimes writes a book. ■

James R. Hansen is professor of history, Auburn University, Auburn, Alabama 36849, USA.



Apollo 11 launched a nation's hopes as the United States raced to the Moon.

NASA

Healthy fats

The Queen of Fats: Why Omega-3s Were Removed from the Western Diet and What We Can Do to Replace Them

by Susan Allport

University of California Press: 2006.

232 pp. \$22.50, £14.95

Asim K. Duttaroy

Two series of polyunsaturated fatty acids, the omega-6 and the omega-3 fatty acids, are essential dietary nutrients. Two major omega-3 fatty acids, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), are increasingly being seen as important modulators of biological pathways that affect growth, development and health. Consumption of these omega-3 fatty acids, which are found in fatty fish and fish oils, has been linked to the low incidence of coronary heart disease in the Inuit people of Greenland. An omega-3 fatty acid with a shorter chain, alpha-linolenic acid (ALA), found in some plant oils and sometimes converted to EPA and DHA, may also be beneficial. Omega-3 fatty acids may protect against cardiovascular disease by several mechanisms, such as lowering blood pressure; reducing plasma triglycerides; lowering the platelet aggregation response, inflammation and arrhythmias; and improving endothelial function, insulin sensitivity and atherosclerotic plaque stability.

Susan Allport discusses the history of using omega-3 fatty acids in her well-written book *The Queen of Fats*. She has done an excellent job in describing the events and the international collaboration involved in discovering the effects of omega-3 fatty acids. She covers almost every aspect of omega-3s, including the importance of an appropriate ratio of omega-6 and omega-3 fatty acids in the diet, which has been biased towards omega-6 fatty acids over the past 50 years, and the consequences for our health of changes in the ratio. There is considerable controversy over how much omega-6 and omega-3 fatty acids are needed in the diet, and which ones.

Many studies have shown a correlation between omega-3 fatty acids and diseases, but translating these findings into recommended amounts for individuals and populations requires careful consideration. Meta-analyses of the effects of omega-3 fatty acids are not conclusive on total mortality, combined cardiovascular events, or cancer. However, some of these meta-analyses failed to take into account many of the pitfalls that must be addressed if a clear picture of the benefits of omega-3 fatty acids is to be obtained.

Allport does not discuss recent meta-analyses of several large studies on fish oils, and fails to mention some of the possible health risks associated with a high intake of fish oils, such as increased bleeding and haemorrhagic



Something fishy: mackerel is a good source of omega-3s.

stroke. In addition, the oxidation of omega-3 fatty acids can increase the level of low-density lipoprotein, reduce glycaemic control in diabetics, and reduce immunocompetence. She could also have elaborated on the significance of another omega-3 fatty acid, DPA, and on adverse health effects in infants and adults of

lipid-soluble toxins (such as dioxins and polychlorinated biphenyl) and of heavy metals in fish oils. I also expected her to say more about the possibility of switching to new sources of

omega-3 fatty acids, such as genetically modified oils, as a way of delivering more highly concentrated sources of omega-3 fatty acids. The advent of algal sources of DHA provides one of the few terrestrial sources of this fatty acid in a concentrated form. All this has been made possible by new technologies and improved processing techniques that ensure the stability and preserve the integrity of these unstable fatty acids.

These omissions aside, however, Allport's book provides an interesting and comprehensive account of the history of omega-3 fatty acids. It not only provides a clearly reasoned case for the benefits of having more omega-3 fatty acids in the diet, but also offers practical advice about how to add these fats to our diet. The book is also well organized, with a

clear contents list, comprehensive notes, time lines and index to enable readers to quickly locate a required topic.

Asim K. Duttaroy is in the Department of Nutrition, University of Oslo, PO Box 1046 Blindern, Oslo N-0316, Norway.

Learning from nature

Biomimetics: Biologically Inspired Technologies

edited by Yoseph Bar-Cohen

CRC Press: 2006. 527 pp. £79.99

Robert W. Cahn

One touch of nature makes the whole world kin —

That all with one consent praise new-born gauds,

Though they are made and moulded of things past,

And give to dust that is a little gilt

More laud than gold o'er-dusted.

Yoseph Bar-Cohen's multi-author volume *Biomimetics*, devoted as it is to a plethora of 'new-born gauds', has examples of gold coated in a variable thickness of dust. In defiance of Ulysses' cynicism in Shakespeare's *Troilus and Cressida*, there is only a little shiny dust on display here, and I suspect that readers will quickly brush it aside.

One of the chapters begins: "The field of

biomimetics encompasses a broad range of topics, generally based on the concept of 'learning from Nature' in areas of Materials Science and Engineering (MSE). This 'learning' may be through inspiration in design, function or a combination of both." Until I read this book, that was my impression of biomimetics too. However, only 5 of the 20 chapters are focused on aspects of MSE, particularly mechanical properties. One of these centres on 'muscles' made from electroactive polymers, and includes some irresistible pictures of a teenage girl arm-wrestling with a synthetic arm.

The rest of the book has a strong emphasis on robotics, both macroscopic and nanoscopic, no doubt prompted by the editor's attachment to the Jet Propulsion Laboratory in Pasadena, with its mission of space and planetary exploration. Synthetic eyes and associated electronics also receive extensive attention. There are two unusual but interesting chapters devoted to design and optimization procedures that imitate the processes of

biological evolution. One of these chapters, on 'genetic algorithms', contains some intriguing examples of this procedure, one of which even incorporates an ingenious analogue of sexual reproduction. Another example involves the optimum locations of a group of post offices to serve a community as economically as possible, and another considers the dynamic balancing of a gas-turbine shaft. Generally, the book contains much control and optimization theory, some of it alarmingly mathematical. It should also be pointed out that most of the optimization procedures discussed make use of computer simulation, rather than physical modelling.

Some chapters include detailed examples of applications of the principles set forth, whereas others, no less stimulating, describe in depth

the relevant features of nature but barely hint at specific applications, such as the chapter on defence and attack mechanisms in biology. Only one (very long) chapter, on the 'mechanization of cognition', seems to be 'gilded dust'; I cannot see that it contributes anything to the theme of the book.

Bar-Cohen himself wrote the substantial opening and closing chapters, and suggests the next stages of research. He also penned the chapter on synthetic muscle, one of his research specialities.

This book differs sharply from earlier treatments of biomimetics, such as R. McNeill Alexander's *Animal Mechanics* (Sidgwick & Jackson, 1968). Stephen Mann's *Biomimetic Materials Chemistry* (VCH, 1996), which I reviewed here ten years ago (*Nature* 382, 684;

1996), focused on materials chemistry. An important subset of this theme is the chemistry and microstructure of high-strength composites, and also spider silk, which clearly exerts great fascination — there is also a chapter on silk in Bar-Cohen's book.

The focus on robotics and on optimization by evolution seems to be unique and constitutes the main claim of Bar-Cohen's volume to widespread attention. Another feature of the book is the extreme care he has taken to obtain the critical opinions of 66 experts, and their ministrations evidently contributed much to the book's clarity of presentation. ■

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

Burning Bush

An exhibition in Australia highlights the country's bushfires.

Colin Martin

Disastrous bushfires, which occur regularly in Australia during the scorching summer months, have scarred the country's collective psyche as well as its landscape. From colonial painter William Strutt's panoramic canvases of bushfires, such as the Black Thursday fires in 1851, to today's television news coverage of widespread forest devastation and burnt-out buildings, images of bushfires abound.

'Fireworks: Tracing the Incendiary in Australian Art' is an exhibition of more than 50 contemporary and historic works that show the impact of bushfires on the country's art. It reveals their continuing potency in stimulating the creativity of Australian artists, who depict both their dramatic pyrotechnic effects and their tragic consequences.

The exhibition includes works by contemporary Aboriginal artists, whose ancestors used fire as a land-management tool over tens of thousands of years. Exploring Australia's east coast in 1770, Captain James Cook described the land as "a continent of smoke" and noted that "we saw smoke by day or fires by night wherever we came". As well as using campfires, Aborigines regularly set fire to areas of woodland and grassland to regenerate plant food for themselves and the animals they hunted, to sustain a viable food chain.

Systematic 'fire-stick farming' released nutrients into the soil and fostered fire-loving plant genera, such as *Banksia* and *Hakea*, which require intense heat to release their seed. By reducing the area of thickly timbered country and creating open grassy woodlands, fire-stick farming limited the spread of fire. Its decline following European



Flames of passion: Tim Storrier's *The Ladder* shows that fire stimulates the imagination.

settlement made it easier for bushfires to spread quickly in southeast Australia.

Many of the exhibited works show the immediate aftermath of bushfires, in charred, still-smouldering landscapes. Others show green undergrowth regenerating among blackened tree trunks, like a phoenix rising from the ashes, hinting at the redemptive power of fire in the Australian landscape. The most dramatic works, however, depict fire as the protagonist. In *The Ladder* by Tim Storrier (shown here), it dominates the foreground. The canvas is covered with shards of flaming light, swirling smoke and heat,

across which a burning ladder has fallen to create a claustrophobic space from which we cannot escape.

The images in 'Fireworks' illuminate how Australian inhabitants have interacted with their environment, from Aborigines to the present-day population. Both scientists and artists, who observe the environment closely in order to depict it, are now encouraging a return to Aboriginal methods of land management to combat bushfires.

'Fireworks' can be seen at the University Art Museum in Brisbane, Queensland, from 19 December until 4 February 2007.

Colin Martin is a London-based writer.

NEWS & VIEWS

SOLID-STATE PHYSICS

Super silicon

Robert J. Cava

Silicon is the archetypal semiconductor, and base material of the microelectronic age. But it turns out that, treated the right way, silicon the semiconductor can become silicon the superconductor.

If someone were to stop me in the street and ask me to name the most important materials on Earth, I would say concrete, steel, glass and silicon. To witness the importance of the first three, just look up from your page or screen. For the last, close your eyes and imagine yourself back in the BS (before-silicon) world of, say, Myrna Loy in *The Thin Man* or Humphrey Bogart in *Casablanca*. One might reasonably argue a preference for the softer focus of those earlier times; but the differences in lifestyle between then and now make it hard to argue against the assertion that silicon has become the technologically most important material of the past 50 years.

It is for this reason that Bustarret and colleagues' report (page 465 of this issue)¹ is such a breakthrough: they have succeeded in turning silicon, the consummate semiconductor, into a superconductor at ambient pressure. Admittedly, the treatment they meted out to silicon to force its conversion ('doping' with high levels of boron) can only be termed abusive, and the temperature at which they measured it (0.3 degrees above absolute zero) frigid. That raises questions as to how useful this turncoat-silicon might be; but its existence is impressive in its own right.

Ironically, the chemical characteristic of silicon that makes it so useful as a semiconductor turns out to be the same characteristic that had stopped it from being made into a superconductor: it generally allows only very small amounts of other elements to be incorporated into its solid form. Silicon can therefore be made extremely pure relatively straightforwardly. Pure silicon is the ultimate semiconductor: it has only a tiny number of free electrons, or positively charged 'holes' (equivalent to electrons missing from the crystal lattice), available to carry electrical current, and so it is effectively electrically insulating. This characteristic can be changed by introducing into the pure silicon small concentrations of atoms that have either one electron more than silicon, such as phosphorus, or one electron fewer, such as boron. This procedure — known as doping — is used to control the number of electrons or holes in carefully patterned regions. Manipulation of these



Window on the world: silicon is arguably the material most central to modern life.

D. HALLINAN/ALAMY

extra charge carriers is the basis of the myriad products based on microelectronics that we know and love.

Over the past decades, doped silicon has proved the perfect platform for testing ideas about the difference, at a fundamental level, between electrically insulating and electrically metallic states of matter (see refs 2,3 and references therein). It is right at the crossover between these two states that superconductivity — when electrons flow without encountering resistance — often occurs.

Here is the gist of what has been found so far. When pure silicon is doped, at low temperatures, with small concentrations of phosphorus — that is, phosphorus atoms replace silicon atoms in the crystal lattice — the extra electron that does not form a crystal bond is localized around the phosphorus nucleus, which has an extra positive charge. Electron and nucleus form a bound state, the one 'orbiting' the other. As the number of phosphorus dopants increases, the nuclei eventually become so close that their electron orbitals overlap, forming one extended orbital. Once that happens, metallic conductivity occurs. This state persists even

down to very low temperatures, despite the disorder caused by the random substitutions of dopant for silicon atoms⁴. The same effect applies to a boron dopant⁵; here, however, the nucleus is negatively charged and the carriers are positively charged.

In the early experiments, silicon was doped with sufficient numbers of electrons and holes to make it a metallic conductor, and the properties of the resulting materials were carefully characterized down to temperatures below 1 kelvin. But superconductivity was never observed.

Bustarret *et al.*¹ adopt a new approach. They dope silicon with boron, to a level well beyond that possible with normal synthesis techniques, by using a pulsed laser to heat a thin silicon film on which a layer of boron-containing gas has been adsorbed. These high-intensity laser blasts cause the boron atoms to invade the silicon, where they freeze in place before the silicon can push them out again. Strange though the resulting structure might be, superconductivity is clearly observed in the treated silicon, turning on at a temperature of 0.3 K.

Performing experiments using such high-powered lasers, and testing materials for superconductivity at such low temperatures, is no small matter. So why bother? The authors are motivated by the possibility that, if silicon could be made superconducting — even under conditions too extreme to be useful in practical devices — the integration of superconducting silicon into the sophisticated world of microelectronics processing might uncover new electronic functions. It will be interesting, for example, to see whether an electron-rich superconductor can be made out of silicon through extreme doping with electron-rich phosphorus or arsenic, rather than hole-rich boron. That would allow the gamut of microelectronics concepts and processing to be applied to superconductors, but is far from an obvious extension of the present work.

So, are Bustarret and colleagues' results¹ just an amusing diversion in the search for new superconductors, or a herald of more and better devices and materials? It's too early to tell. The main thrusts in the search for new

superconducting materials nowadays are towards exotic systems in which magnetism can be transformed into superconductivity by changing a carefully controlled experimental parameter, and towards materials based on metallic elements that have high superconducting transition temperatures and are easy to process. Such a material could change the way electrical current is carried. I remain convinced that we could wake up any morning to the announcement that someone, somewhere, has found it. If that superconductor is made by doping concrete, I'll know it's time for me to retire. ■

Robert J. Cava is in the Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.
e-mail: rcava@princeton.edu

1. Bustarret, E. *et al.* *Nature* **444**, 465–468 (2006).
2. Lee, P. A. & Ramakrishnan, T. V. *Rev. Mod. Phys.* **57**, 287–337 (1985).
3. Thomas, G. A. *Phil. Mag.* **B52**, 479–498 (1985).
4. Rosenbaum, T. F. *et al.* *Phys. Rev. Lett.* **45**, 1723–1726 (1980).
5. Dai, P., Zhang, Y. & Sarachick, M. P. *Phys. Rev. Lett.* **66**, 1914–1916 (1991).

HUMAN GENOMICS

In search of normality

Kevin V. Shianna and Huntington F. Willard

The first map of copy-number variation in the human genome has been created. It is now feasible to examine the role of such genome variation in disease and to explore in depth the extent of 'normal' variability.

The human genome contains many forms of genetic variation. The most plentiful are the millions of single base-pair changes in the DNA code that were identified in the course of determining the human genome sequence, and then more systematically through the International HapMap Project¹. These so-called single nucleotide polymorphisms (SNPs) distinguish any two unrelated copies of the genome. They account for the long-hypothesized, evolutionarily 'neutral' forms of widespread genetic variation that mark diversity within our species, as well as mutations, both rare and common, that account for or contribute to disease.

Less expected have been variations in the copy number of sequence elements — that is, variation in the number of deleted or duplicated versions of segments of the genome that result in a range of the number of copies (instead of the usual two) among apparently 'normal' members of the population². Several studies have described the prevalence of common deletion polymorphisms in the human genome^{3,4}. On page 444 of this issue, Redon *et al.*⁵ now present results of a global genome-wide screen looking for all types of copy-number variants (CNVs) using several hundred reference samples from four human populations. They document nearly 1,500

variable regions, covering a remarkable 12% of the human genome and including hundreds of genes and other functional elements whose copy number differs, sometimes dramatically, among us. The data suggest that the greatest source of genetic diversity in our species lies not in millions of SNPs, but rather in larger segments of the genome whose presence or absence calls into question what exactly is a 'normal' human genome.

To detect CNVs, Redon *et al.* used two complementary genome-wide technologies. The first was a genotyping approach in which some 500,000 SNPs were assayed, looking for stretches of adjacent SNPs that displayed atypical ratios of the expected two versions (called alleles) of a given SNP. The second involved comparing each sample with a reference standard, and looking for systematic differences in intensity among a set of more than 26,000 large-insert cloned segments that span nearly all of the currently sequenced portion of the genome.

Combining these approaches provided coverage adequate to detect most forms of CNVs. In total, 1,447 CNVs were identified across the 270 HapMap samples. The estimated average length of CNV regions per genome analysed was more than 20 million base pairs, representing some 5- to 10-fold more variation between

any two randomly chosen genomes than suggested previously by studying SNPs alone. More than half of the CNVs that were identified overlap known annotated genes in the genome. So it is likely that CNVs play a role in so-called complex diseases, in which multiple genes and/or gene–environment interactions are involved.

Mechanistically, how might copy-number variation be involved with complex disease? When deletions or duplications are present within a gene or its regulatory region, there is a reasonable chance that there will be an imbalance in the appropriate level of RNA and thus protein production from that gene. For genes and pathways in which the amount of a functional product produced is critical, it seems likely that CNVs could underscore variation in susceptibility to disease. Classically, variation in the copy number of the globin genes was shown to be responsible for various disorders of haemoglobin, such as the α -thalassaemias⁶. More recently, variable copy number of the *CCL3L1* gene was reported to be associated with increased resistance to infection by HIV⁷.

Many genome-wide studies are currently under way that aim to find SNPs associated with complex disease. These studies in effect look for disease- and population-specific changes in the frequencies of SNP alleles, using arrays containing 'tagging SNPs' that act as proxies for other closely associated SNPs that are inherited together as a block. The likely involvement of CNVs in complex disease, however, raises the question of whether the many CNVs reported by Redon *et al.*⁵ can also be detected by association with one of these tagging SNPs. The answer seems to be both yes and no. Some CNVs may be associated with their neighbouring SNPs over time, but others may be of newer origin and their presence or absence may not be accurately tagged. Thus, densely spaced SNPs will probably be a prerequisite for the most efficient use of CNVs in genome-wide SNP-based association studies.

Given the limited set of reference samples assayed, the 1,500 CNVs reported by Redon *et al.* are probably the tip of the iceberg. As the results and the raw data from the first wave of genome-wide association studies become available, it will be essential to catalogue the full range of human CNVs. A complete map of CNVs in global populations will be necessary before we can fully understand which of the variants have clinical or other consequences, and which are, in fact, within the extremes of what we consider 'normal'.

More than a hundred years ago — even before the rediscovery of Mendel's laws of inheritance, and well before an awareness of DNA and genomes — the physician Sir Archibald Garrod first shed light on what he termed "chemical individuality" to refer to biochemical variants that shape the intricacies of metabolism in different individuals⁸.

As he observed⁹ so presciently, “the existence of chemical individuality follows of necessity from that of chemical specificity, but we should expect the differences between individuals to be still more subtle and difficult of detection.” Our current view of “genomic individuality” at the level of SNPs, CNVs and chromosomal variants indeed extends his view in ways “more subtle and difficult of detection”. The stage is set for global studies to explore anew, as Garrod once did, the clinical significance of human variation.

Kevin V. Shianna and Huntington F. Willard are at the Institute for Genome Sciences and Policy,

Duke University, Durham, North Carolina 27708, USA.

e-mails: k.shianna@duke.edu;
hunt.willard@duke.edu

1. International HapMap Consortium *Nature* **437**, 1299–1320 (2005).
2. Feuk, L. *et al.* *Nature Rev. Genet.* **7**, 85–97 (2006).
3. Hinds, D. A. *et al.* *Nature Genet.* **38**, 82–85 (2005).
4. McCarroll, S. A. *et al.* *Nature Genet.* **38**, 86–92 (2005).
5. Redon, R. *et al.* *Nature* **444**, 444–454 (2006).
6. Weatherall, D. J. *Am. J. Hum. Genet.* **74**, 385–392 (2004).
7. Gonzalez, E. *et al.* *Science* **307**, 1434–1440 (2005).
8. Garrod, A. E. *Mol. Med.* **2**, 274–282 (1996; reprint of original 1902 paper).
9. Garrod, A. *Inborn Errors of Metabolism* 2nd edn (Oxford Univ. Press, 1923).

CELL BIOLOGY

Infectious Alzheimer's disease?

Roland Riek

Accumulation of organized, self-polymerizing protein aggregates is a hallmark of Alzheimer's disease and infectious prion diseases. The similarities between these conditions may be even closer than that.

Amyloid fibrils are malicious. These insoluble, highly organized protein aggregates are associated with devastating disorders such as Alzheimer's and Parkinson's diseases, type II diabetes and the prion (proteinaceous infectious particle) diseases that include Creutzfeldt–Jakob and mad cow diseases¹. The aggregates are toxic, highly stable and can jump-start self-polymerization by recruiting their normal, soluble protein counterpart². Amyloid self-polymerization is also the basis of the ‘protein-only’ hypothesis for the mechanism of prion infectivity³: the infectious prion conformation replicates itself in a host by catching the ‘benign’ host prion protein and forcing it into the infectious conformation. The nature of this two-component system means that the conformations of both the prion agent and the host prion protein influence the characteristics (phenotype) of the resulting prion strains, such as age of onset of the disease in the host or brain localization of the amyloid⁴.

If a prion amyloid is infectious, can other amyloid disorders such as Alzheimer's disease also be infectious? This is the provocative question raised by Meyer-Luehmann *et al.*⁵ in *Science*. They report data showing that amyloids of the amyloid- β (A β) peptide that is associated with Alzheimer's disease behave like an infectious agent when injected into the brain of a mouse model of Alzheimer's disease, generating phenotypes that depend on both the host and the agent.

Aggregation of the A β peptide into amyloid fibrils is a central process in Alzheimer's and other diseases¹ (Fig. 1). *In vitro* studies show that amyloid formation is affected by both elapsed time and A β concentration, but that it can be rapidly initiated by the addition

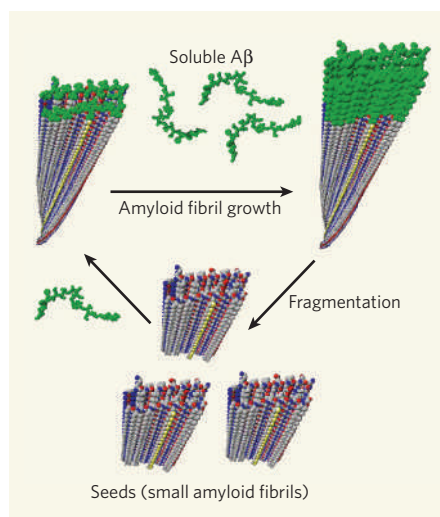


Figure 1 | Replication of amyloid fibrils. These fibrils are associated with Alzheimer's and other diseases, as well as prion disorders, and grow by recruiting their normal, soluble protein counterpart. Fragmentation of the fibrils produces seeds, which in turn prompt further fibril formation. Meyer-Luehmann *et al.*⁵ find that, when amyloid- β (A β) peptide is injected into the brains of mice with an Alzheimer's-like disease, seeded aggregation occurs.

of a ‘seed’ consisting of A β amyloids².

Meyer-Luehmann *et al.*⁵ now provide *in vivo* evidence of such a time- and concentration-dependent seeded aggregation of A β . Their experiments used brain extracts from deceased Alzheimer's patients or from aged mice genetically predisposed to develop an Alzheimer's-like disease (transgenic Alzheimer's mice), or synthetic A β . The extracts were injected into the brains of young,

presymptomatic transgenic Alzheimer's mice. In all instances, the authors observed seeding activity (although with reduced efficiency in the case of synthetic amyloid). A β peptides seem to be responsible for the seeding activity because this activity was lost if injected extracts were used in which either the A β conformation had been destroyed or A β removed. These studies further strengthen observations of *in vivo* seeding of A β in transgenic mice⁶ and in non-human primates⁷, suggesting a mechanism reminiscent of prion infectivity. Furthermore, the pathology of A β formation is governed by both the injected agent and the host, features typically observed for prion strains.

Meyer-Luehmann *et al.*⁵ used two distinct transgenic mouse models of Alzheimer's disease: APP23 mice and APPPS1 mice. APP23 mice overproduce mainly the 40-amino-acid A β peptide, which results in the deposition of predominantly diffuse and filamentous A β in aged animals. APPPS1 mice overproduce mainly the 42-amino-acid A β , and show a compact (punctate) deposition of A β with age. Injection of brain extracts from aged APPPS1 mice into young APP23 hosts consistently induced punctate A β deposition that was mainly confined to one part of the brain (the subgranular layer of the hippocampus). In contrast, extracts from aged APP23 mice injected into APP23 hosts yielded primarily diffuse and filamentous lesions, with substantial diffuse A β deposition. When the experiments were applied to the APPPS1 host, deposition induced by the APPPS1 extract was even more punctate, whereas that induced by the APP23 extract was a mixture of the filamentous and punctate A β forms.

What factors might make an amyloid infectious? According to a mathematical model⁸ validated by experiments in three strains of yeast prion, the generation of a prion and its phenotype depends on several factors: the amount of amyloid; the conformation-dependent growth rate of the amyloid; the division rate of the amyloid into seeds; and the concentration of the soluble protein counterpart.

On this account, manipulation of any of these variables can transform any amyloid into a prion. For example, an increase of soluble amyloid protein in the host might be sufficient to transform an amyloid into a prion. In contrast, the formation of long fibrils decreases the ratio of the number of seeds to amyloid protein, reducing the seeding capacity. The amyloid conformation determines its growth and division rates, thereby influencing the titre of infectivity and the prion phenotype. Indeed, Meyer-Luehmann *et al.*⁵ observed induction of A β formation by A β injection only in A β -overproducing transgenic mice and not in normal mice. Injection of long, synthetic A β fibrils resulted in a low seeding capacity, and the different amyloid phenotypes described might be attributed to the distinct properties and three-dimensional structures of amyloids consisting of either the 40- or the 42-amino-acid A β ^{9,10}.

It might sound shocking that Alzheimer's disease, and possibly other amyloid diseases, can be infectious under certain circumstances (in this case⁵, intracerebral injection into A β -overproducing mice). According to the nucleation–polymerization model of amyloids, however, it is to be expected. So where do we go from here?

The experimental evidence^{5–7} and widespread incidence of amyloid diseases highlight the need for basic, clinical and epidemiological studies into the possible transmission of amyloid diseases between humans. In the United States alone, 4.5 million people suffer from Alzheimer's disease, and 20 million from type II diabetes; in particular, we need to find out whether seeds can be transmitted and can accelerate the onset of disease in humans.

For those involved in basic research, there are plenty of other issues to tackle. Is the use of mice that overexpress prion protein still adequate to study transmissible spongiform encephalopathies, given that overexpression might mask the real properties of the agent¹¹? How many strains of Alzheimer's disease are there, and how well are they reflected in the

established transgenic mouse strains? If there are several strains — and so several amyloid conformations — how relevant are the structural studies of amyloids that are generated *in vitro*? Amyloids are certainly malicious in their association with human disease. They are also highly challenging in the increasing number of questions they pose for researchers.

Roland Riek is in the Department of Structural Biology, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.
e-mail: riek@salk.edu

1. Selkoe, D. J. *Nature* **426**, 900–904 (2003).
2. Harper, J. D. & Lansbury, P. T. *Annu. Rev. Biochem.* **66**, 385–407 (1997).
3. Prusiner, S. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
4. Weissman, C. *Nature Rev. Microbiol.* **2**, 861–871 (2004).
5. Meyer-Luehmann, M. et al. *Science* **313**, 1781–1784 (2006).
6. Kane, M. D. et al. *J. Neurosci.* **20**, 3606–3611 (2000).
7. Ridley, R. M. et al. *J. Neural Transm.* **113**, 1243–1251 (2006).
8. Tanaka, M. et al. *Nature* **442**, 585–589 (2006).
9. Petkova, A. T. et al. *Biochemistry* **45**, 498–512 (2006).
10. Luhrs, T. et al. *Proc. Natl Acad. Sci. USA* **102**, 17342–17347 (2005).
11. Aguzzi, A. & Heikenwalder, M. *Nature Rev. Microbiol.* **4**, 765–775 (2006).

tools have been developed for transforming two-dimensional optical spectra into molecular structures. These tools have been successfully applied to provide information about a variety of processes, such as the making and breaking of hydrogen bonds between solute and solvent molecules^{2–5}, energy transfers in a photosynthetic light-harvesting protein⁶, and changes in distance between oscillating groups found in biomolecules^{1,7}.

Throughout these theoretical and experimental efforts, a recurrent theme has emerged: the effects of interactions between molecular groups are tiny. One-dimensional optical spectroscopy is used to measure the so-called primary properties of vibrating groups in molecules, such as their oscillation frequencies and dipole strengths. But interactions between different oscillators — such as couplings between vibrational or electronic oscillations — are known as secondary properties, and are up to 1,000 times smaller than the primary properties. Secondary properties cannot be selectively measured by most conventional spectroscopic techniques, because they are usually hidden under the dominant primary features. But it is the secondary properties that are of most interest when studying the way that a biomolecule folds and unfolds. The precise three-dimensional structure of a molecule can have a profound effect on the secondary properties of vibrating groups within that molecule, because these are often proportional to $1/R^n$ (where R is the distance between two oscillators and n is greater than or equal to 3).

Fortunately, two-dimensional optical spectroscopy is a superb tool for measuring the secondary properties of molecular oscillators, using a process known as differencing. In this technique, one spectrum is subtracted from another that has been obtained under alternative conditions, so that the observed differences can be correlated with secondary properties.

SPECTROSCOPY

Molecular motion pictures

Minhaeng Cho

Molecules in solution change their conformations at ultrafast speeds, so that no method has been able to record the process. This looks set to change, as infrared spectroscopy rises to the challenge.

The science of high-speed photography was created by Eadweard Muybridge in 1887, when he published his famous photographic sequence of a galloping racehorse. Scientists have since sought analogous microscopic and spectroscopic techniques to enable them to watch the motion of biological molecules in real time. As reported on page 469 of this issue¹, Hamm and co-workers have transferred the principles of high-speed photography to two-dimensional infrared spectroscopy. In this way, events that occur over a few hundred picoseconds can be studied by taking a series of snapshots — that is, two-dimensional infrared spectra — of the molecules concerned.

Hamm and colleagues' technique is not just an ultra-high-speed version of Muybridge's method that uses infrared radiation; the smaller scale requires notable differences. The subject of the authors' study¹ is a loop of four amino acids, known as a cyclic tetrapeptide, that is a billion times smaller than a horse. More pertinently, this tetrapeptide is 100 to 1,000 times smaller than the wavelengths of infrared or visible radiation, so it is impossible to simply take a photograph of it. Instead, the authors recorded a spectrum in which the absorp-

tion bands are associated with polarization changes induced by interactions of infrared radiation with the molecule. Many theoretical

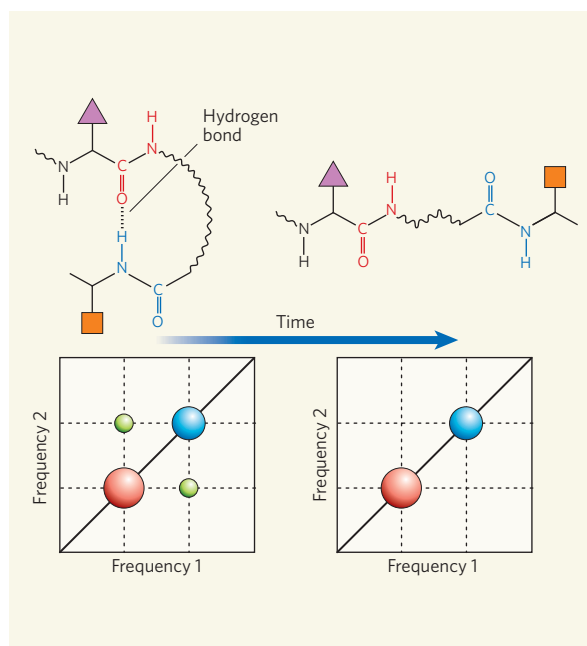


Figure 1 | Spectroscopy of changes in molecular structure. The peaks on the diagonal line of this typical two-dimensional infrared spectrum are caused by vibrations of the chemical groups in red and blue in the structures above (the square and the triangle represent amino-acid side groups). The cross-peaks in green are caused by the coupling of these vibrations. As the molecule unfolds, the length of the hydrogen bond increases and the vibrational coupling decreases, so that the cross-peaks become less intense. The cross-peaks disappear when the hydrogen bond is broken. By examining the amplitudes of cross-peaks from such spectra, Hamm and colleagues¹ followed the breaking of a hydrogen bond, and so the structural evolution of a small molecule, with time.



50 YEARS AGO

The possible changes, especially perhaps in the emotional sphere, brought on by the stimulating action of salt are, of course, entirely a matter of speculation. The greater responsiveness of people, if they were so stimulated, could have helped throughout the ages in the accumulation of knowledge... There is no question that there is a sound basis for the prescribing of low-salt diets in many diseases, particularly those involving the circulatory system. When it comes to normal people, however, recommendations are infinitely more difficult. It is certainly true that the chemistry of the body does not require the addition of salt to our food. The physician, however, is not primarily interested in the mere metabolic processes but in the general welfare of his patients, and he should consider that the quickened pace of a more complicated society demands persons with a heightened responsiveness. Salt may be one of the ingredients producing this effect.

From *Nature* 24 November 1956.

100 YEARS AGO

The author [of *The Human Epic*] begins with the evolution of the earth and the origin of life, and strives to show the changes undergone by the inorganic world and the gradual appearance of lowly marine beings in the Cambrian and Silurian seas. Of poetic fancies the author nothing lacks, but of natural history lore his stock is meagre:—

"Much fear I him who armed with
claws and quills
Steals stealthily along the weedy mire.
I dread the shape who bears the
bristling gills
Which seem with rage and venom to
respire.
But chiefly do I fear the lobster dire.
Four claws he wears, his quarry to assail,
Two spears he brandishes to wreak
his ire,
Invulnerable gleams his quilted mail.
O'er such stupendous foe nought living
can prevail"

We are quite at a loss to fit the author's description with any Silurian, or, indeed, any other fossil arthropod!

From *Nature* 22 November 1906.

The transient two-dimensional infrared spectroscopy used by Hamm and colleagues¹ is actually a double differencing technique. To remove the obscuring layer of primary contributions, they recorded a 'pump-probe' spectrum; this is the difference between a spectrum obtained using a pump pulse of infrared (which excites the molecule) and a spectrum obtained using a probe pulse (which records the temporal evolution of the excited state). To obtain an even sharper glimpse of the temporal changes in secondary properties, differences between spectra recorded at different times are also calculated.

The authors used their technique to watch a biomolecule unfolding in solution, by observing the breakage of a hydrogen bond. They prepared a cyclic tetrapeptide containing a weak chemical bond known as a disulphide bridge, that holds the molecule in a folded conformation. They then used a picosecond pulse of ultraviolet light to break the disulphide bridge. This triggering reaction allowed the unrestricted molecule to unfold, and the authors recorded its movement as a series of two-dimensional infrared spectra using femtosecond pulses of infrared. One could compare Hamm's team to film directors, cueing the molecular action with a burst of ultraviolet and using infrared spectra as their movie frames; the shutter speed of their 'camera' is the duration of an infrared pulse, and the exposure time for each frame is the period between pairs of pulses.

By subtracting a reference spectrum from the transient ones, the authors revealed the secondary properties associated with the molecule's motion as cross-peaks in the spectra (Fig. 1). They identified four oscillators in the

tetrapeptide, and with the help of computer simulations they were able to describe the temporal evolution of the cross-peaks' amplitude in terms of the time-dependent changes of length for a particular hydrogen bond. In other words, they were able to 'watch' the dynamics of the hydrogen bond as it stretched and broke.

Two-dimensional infrared spectroscopy is currently limited to studying certain types of oscillator in proteins, as only a narrow range of oscillation frequencies is visible using this technique. Furthermore, it faces increasing competition from an analogous technique in which optical spectroscopy is used to study coupling between electrons⁶. Although two-dimensional infrared spectroscopy has not yet burgeoned into a systematic discipline, its applications — such as studying the ultrafast dynamics of biomolecular and chemical reactions, or of protein interactions with other molecules — may be waiting just around the corner. With this approach, Hamm and colleagues have shown that the movements of participants in molecular dramas can be recorded in vivid detail. ■

Minhaeng Cho is in the Department of Chemistry, Korea University, Seoul 136-701, South Korea.

e-mail: mcho@korea.ac.kr

1. Kolano, C., Helbing, J., Kozinski, M., Sander, W. & Hamm, P. *Nature* **444**, 469–472 (2006).
2. Kwac, K. & Cho, M. *J. Chem. Phys.* **120**, 1477–1490 (2004).
3. Kim, Y. S. & Hochstrasser, R. M. *Proc. Natl Acad. Sci. USA* **102**, 11185–11190 (2005).
4. Zeng, J. *et al. Science* **309**, 1338–1343 (2005).
5. Cowan, M. L. *et al. Nature* **434**, 199–202 (2005).
6. Brixner, T. *et al. Nature* **434**, 625–628 (2005).
7. Chung, H. S. & Tokmakoff, A. *Proc. Natl Acad. Sci. USA* **102**, 612–617 (2005).

MEDICINE

Blastomeres and stem cells

Joe Leigh Simpson

Generating human stem cells from a single cell recovered during preimplantation genetic diagnosis does not, in principle, harm the embryo. Can the approach be used in assisted reproductive technology programmes?

Elsewhere in this issue, Klimanskaya *et al.*¹ report an exciting advance for the field of human embryonic stem (hES) cell biology. Readers may have a sense of déjà vu, however, for this paper first appeared as an online publication² on 23 August. It was the subject of controversy, partly because of confusion over certain points. It now appears both in print (see page 481) and on the *Nature* website with an explanatory addendum.

The promise of regenerative medicine through stem-cell therapy is intoxicating. Stem-cell lines are pluripotent (that is, they can develop into various cell types); they are self-renewable; and they can differentiate into functional cells. Scientific hurdles remain,

principally in inducing nascent stem cells into the desired differentiation pathway. But the potential for treating intractable human disease, by using stem cells to repair damaged tissue, is real. This, however, is a topic that is as much about ethics and politics as about science.

The approach of Klimanskaya *et al.*¹ offers a possible route around the objection that the creation of hES cells involves destroying the embryo from which they arise. The authors propose that it could be applied as part of the established technique of preimplantation genetic diagnosis (PGD), which is used to identify genetic defects in embryos created through *in vitro* fertilization before they are implanted in the uterus.

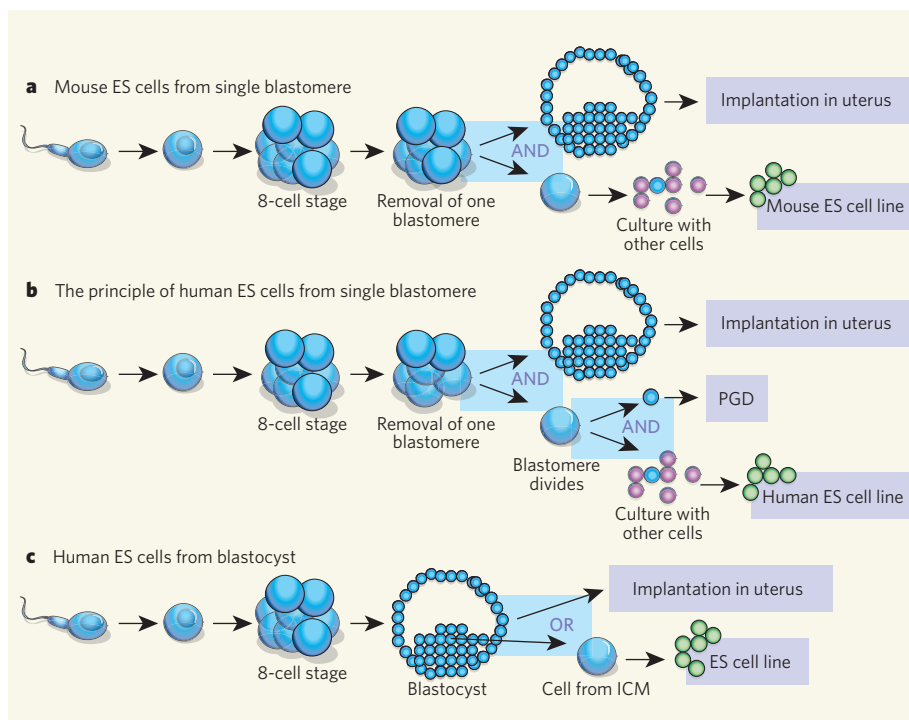


Figure 1 | Production of embryonic stem-cell lines. **a**, The experiment with mice reported earlier this year³. A blastomere removed at the 8-cell stage was used to produce stem cells; the remaining 7-cell-embryo was allowed to develop further, was implanted in the uterus and went on to form an intact embryo. **b**, The idea behind using human preimplantation genetic diagnosis (PGD) for hES cell production. The biopsy is the same as in **a**, but the single blastomere is allowed to divide. One cell is used for PGD and the other for hES cell production. This scheme shows the 7-cell-embryo being allowed to develop further, as it would in practice, but in the experiments of Klimanskaya *et al.*¹ no embryos were produced because all blastomeres were removed. **c**, Production of hES cells from the inner cell mass (ICM) at the blastocyst stage. This approach is the best developed but involves destroying the blastocyst.

Their procedure¹ was as follows. During the early stages of embryo development, the fertilized egg starts to divide into cells called blastomeres. Klimanskaya *et al.* separated individual blastomeres from human embryos at the 8-cell stage, using a micromanipulation method widely used in PGD. Multiple blastomeres from a given embryo were singly placed in individual wells within a flask and allowed to divide. Approximately half did so, and were subject to further culture, as described in the paper. The original cohort of 91 blastomeres yielded 19 ES-cell-like outgrowths and two stable hES cell lines. This work with human blastomeres follows a demonstration by the same group that ES cells can be derived from single mouse blastomeres (Fig. 1a)³. In these earlier mouse experiments, an intact viable embryo developed that consisted of the seven remaining blastomeres; by contrast, in the work with human cells, multiple blastomeres were taken from the 8-cell stage and no embryos were allowed to remain in culture. This was a source of confusion in the earlier online publication².

The main question raised by Klimanskaya *et al.*¹ is whether a single human blastomere, already required for PGD, could be allowed to divide prior to genetic analysis: one daughter cell would then be used for clinical diagnosis, whereas the other would be used to derive an

hES cell line. The remaining 7-cell ball of blastomeres would be implanted in the uterus to develop normally, as is PGD practice (Fig. 1b). The advantage of this procedure would be in creating additional hES cell lines. Parents could possibly bank hES cells from the child from which the blastomere came, so their child could benefit from patient-derived hES-cell therapies if they come to fruition. Furthermore, banking of large numbers of hES cells could be a way of providing immunologically compatible cell lines for a large proportion of the population.

Existing hES cell lines are derived from a later stage of embryo development — the preimplantation blastocyst, 5 to 6 days after conception (Fig. 1c). A blastocyst is needed because its inner cell mass (ICM) is necessary to generate stem cells; outer cells (trophoblasts) differentiate into the placenta. All hES cell lines 'approved' by the US National Institutes of Health are derived from an ICM, as, presumably, are the less publicized, unapproved hES cell lines. Yet there are deep divisions among both the public and politicians over whether it is ethical to use an ICM to create hES cells, even for therapeutic purposes. This is because of the need to destroy the embryo from which the ICM is obtained.

To obviate this (and perhaps help lift the embargo on federal research funding in the United States), various other alternatives

have been proposed, as described in Box 1 (overleaf). The current contribution¹ offers another route.

In PGD, the most common technique involves removing a single blastomere from the 8-cell (3-day) embryo. The blastomere is analysed to detect genetic abnormalities; the 7-cell embryos that have no abnormalities are then transferred to the mother's uterus for implantation and pregnancy. A variant involves analysis of the first and second polar bodies, products of early oocyte division, whose genotype can predict that of the oocyte that will contribute the female lineage⁴. In both approaches, embryo transfer takes place 24–48 hours after removal of a blastomere or polar body, although transfer can be delayed until the blastocyst stage (day 5 or day 6). The main point, however, is that there is a limited time window for blastomere removal, genetic analysis and transfer.

Is the procedure proposed by Klimanskaya *et al.*¹ feasible in this context? Certainly there are ample cases of PGD, which is an increasingly attractive complement to traditional prenatal genetic diagnosis and is often preferred for identifying chromosomal abnormalities. Use of PGD decreases miscarriages, increases implantation rates and almost certainly increases live-birth rates in properly selected cases^{5,6}.

For blastomere-derived hES cells to become a preferred option, all concerned would need assurance that deferring genetic analysis by allowing the blastomere to divide would decrease neither the already low pregnancy rate of *in vitro* fertilization (only 25–30%) nor diagnostic accuracy. Is the likelihood of providing accurate PGD diagnosis compromised as a result of waiting for cell division, potentially risking cell degeneration? In fact, only half the blastomeres divided in the current report¹.

Safety is another concern. PGD apparently does not increase the frequency of birth defects in babies, but more than the usual manipulation might be necessary in generating blastomere-derived ES cells. Sequential PGD biopsies (first polar body, second polar body, 8-cell embryo) surprisingly seem not to diminish pregnancy rates⁴. If diagnosis is compromised by cell division, however, the temptation might arise to remove two blastomeres, one for PGD and one for hES cells. This surely would adversely affect embryo viability.

Potentially more problematic are the consequences of certain technical nuances used in Klimanskaya and colleagues' method¹. Multiple isolated blastomeres from a single embryo were cultured together, presumably to derive benefits from cell–cell interactions. Because only one blastomere would be removed in clinical PGD, the authors propose that the blastomere should be co-cultured with its embryo; the embryo would later be transferred into the uterus to produce a pregnancy. The safety of this approach may depend on whether deleterious effects arise as a result of embryo culture, for allowing the blastomere to divide prior to genetic analysis means that the embryo

has to remain longer *in vitro*. Perturbations of genomic imprinting due to the possibly longer time needed in culture are the underlying scientific concern. In imprinting, the expression of certain genes depends on whether they come from the mother or father; culture media may or may not mimic the milieu of the human reproductive tract in directing correct parental expression.

Despite these reservations, the derivation of hES cell lines from a single human blastomere using clinical PGD is realistic, and offers an attractive way out of an ethical conundrum. As Irving Weissman remarked in a previous News & Views article⁷, some politicians are already advocating that generating hES cells from the ICM should cease. Research would be redirected towards blastomere-derived or 'inhibited' embryonic lines (Box 1) — if indeed stem cells of embryonic origin are not abandoned altogether in favour of cell lines solely of adult origin, should that more demanding approach prove feasible. Yet efficacy of the alternatives to ICM-derived lines remains unclear. The reality is that ICM-derived hES cell lines exist, and in increasing number. The same cannot be said about the alternatives: research with ICM-derived hES cells should not be held in abeyance.

In the long term, will the method of deriving hES cells matter? Perhaps not. If we achieve a successful example of human stem-cell therapy

Box 1 | Other prospects for producing stem cells

● **'Inhibited' cell lines.** This was suggested by work⁹ in which a gene inhibiting expression of another gene, *Cdx2*, was inserted into an embryonic donor cell; *Cdx2* is required for establishing the trophectoderm necessary for placental formation. After pluripotent stem cells developed, the inhibiting gene was removed. This allowed differentiation into a stem-cell line, but too late for potential implantation. The approach should mollify those fearful of reproductive cloning, but it does not address the fundamental concern over disturbing the inner cell mass.

● **Stem cells from adults.** Although this would be ideal, efficacy remains highly speculative.

● **Targeted, patient-specific hES cells.** This process is even more challenging, and requires a process called somatic cell nuclear transfer. The nucleus of an immature egg, an oocyte, is removed and in its place is inserted a nucleus of an adult cell from the patient in question. The ES cell line generated will thus have the patient's genotype, and immunological incompatibility should not be a problem.

As discussed in the main text, however, the derivation of hES cells from the inner cell mass (Fig. 1c) remains the most successful route. **J.L.S.**

with ICM-derived cells, the controversy over the provenance of the cells is likely to dissipate in large part. After all, opprobrium was initially heaped on advocates of prenatal genetic diagnosis in the 1960s, of *in vitro* fertilization in the 1970s, and of PGD in the 1990s. Each success was followed by widespread acceptance of these procedures⁸. Time and public opinion move on.

Joe Leigh Simpson is in the Departments of Obstetrics and Gynecology, and Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

e-mail: jsimpson@bcm.tmc.edu

1. Klimanskaya, I., Chung, Y., Becker, S., Lu, S.-J. & Lanza, R. *Nature* **444**, 481–485 (2006).
2. Klimanskaya, I., Chung, Y., Becker, S., Lu, S.-J. & Lanza, R. *Nature* doi:10.1038/nature05142 (2006).
3. Chung, Y. *et al. Nature* **439**, 216–219 (2006).
4. Verlinsky, Y. & Kuliev, A. *Practical Preimplantation Genetic Diagnosis* (Springer, New York, 2005).
5. Munné, S. *et al. Reprod. BioMed. Online* **7**, 91–97 (2003).
6. Gianaroli, L. *et al. Reprod. BioMed. Online* **10**, 633–640 (2005).
7. Weissman, I. L. *Nature* **439**, 145–146 (2006).
8. Simpson, J. L. *Reprod. Biomed. Online* www.rbmonline.com/Article/2466 (2006).
9. Meissner, A. & Jaenisch, R. *Nature* **439**, 212–215 (2006).

MOLECULAR BIOLOGY

Triggering positive competition

Ada Yonath

In most bacteria, a molecule known as trigger factor prevents misfolding of newly made proteins emerging from their ribosome factory. The dynamic action of this molecule has been followed using fluorescence spectroscopy.

Proteins have specific structures designed for their tasks, the fold of each protein being dictated by its amino-acid sequence. Within the cell, protein folding occurs as the protein is being made by a multi-subunit complex called the ribosome. Before the protein is long enough to acquire its final fold, one end of it emerges from the protective exit tunnel of the ribosome into the crowded cellular environment. To promote efficient folding under these unfavourable conditions, all cells contain molecular 'chaperone' proteins. The various effects exerted by these chaperones are only partly understood, but some of them have the task of preventing the aggregation and misfolding of the emerging proteins¹, mostly by transiently masking 'sticky' hydrophobic surfaces. These sticky patches will generally become buried inside the mature protein, but may be exposed on the elongating polypeptide chains.

In eubacteria, this task is performed by a ribosome-binding chaperone called trigger

factor — eubacteria being the largest group of bacterial species, and the only subgroup that contains trigger factor. On page 455 of this issue, Kaiser *et al.*² describe how they have watched trigger factor at work, providing insight into its protective mechanism (Fig. 1, overleaf).

Hints of how trigger factor might operate have come from static binding experiments and crystal structures^{3–5}. By contrast, Kaiser *et al.* used an ingenious fluorescence spectroscopy technique to follow trigger factor dynamically as it interacted with the ribosome and a nascent polypeptide. Consistent with the crystal structures of physiologically meaningful complexes^{4,5}, they found that when trigger factor binds to the ribosome it undergoes conformational rearrangements and acquires an 'open' activated form that can adhere to hydrophobic patches on the polypeptide. Furthermore, in cases when many hydrophobic segments are exposed, the activated trigger-factor molecule may remain associated with the elongating

nascent chain even after the polypeptide has left the ribosome, permitting another molecule of trigger factor to dock on the ribosome. So, until it reaches the next chaperone, the emerging protein may be protected by one or more trigger-factor molecules, with the segments between them probably being protected from degradation because they are partially folded or because of steric hindrance.

This dynamic view is consistent with the fact that all cells contain a significant excess of trigger factor over the number of ribosomes, even though in the cytosol trigger factor binds to almost all active ribosomes at a 1:1 ratio⁶. This means that there is a continuous supply of trigger factor to protect a nascent chain. In the cytosol, unbound trigger-factor molecules exist mainly in pairs; however, they bind to the ribosome singly, contrary to a previous suggestion⁷. The equilibrium between pairs and single molecules might be used to control the amount of readily available trigger-factor molecules for

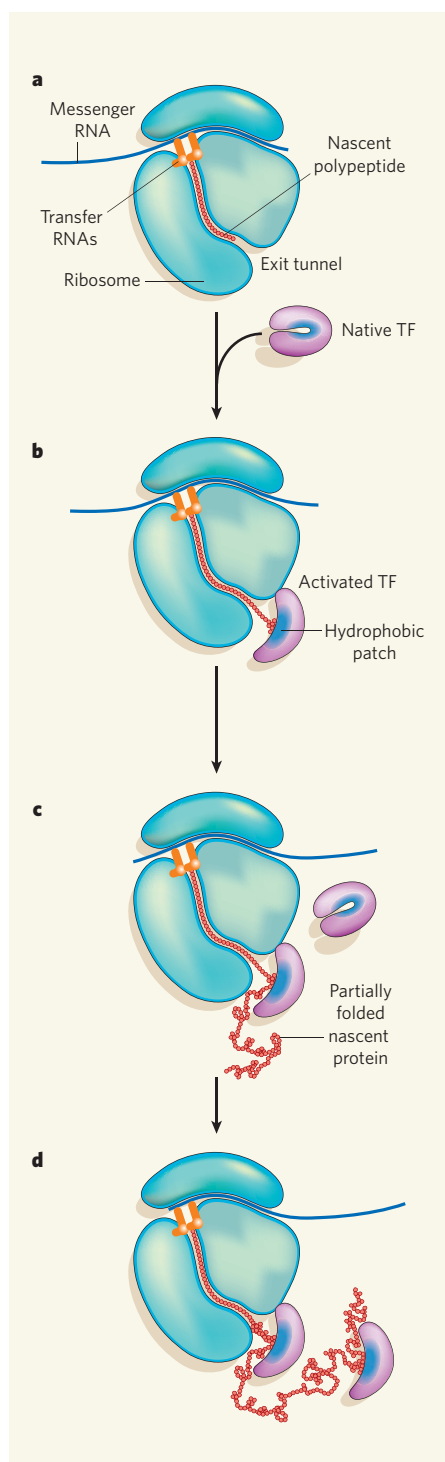


Figure 1 | Trigger factor and protein production by eubacteria². **a**, The code in messenger RNA is read by the ribosome, with transfer RNAs bound to the correct amino acids to form the nascent polypeptide chain. **b**, As the polypeptide emerges from the exit tunnel, trigger factor (TF) in the cytosol is activated and changes conformation; it docks on the ribosome, produces a partially closed shelter, and binds hydrophobic patches of emerging protein chain, preventing misfolding. **c, d**, Activated TF can remain associated with the nascent protein even after it has left the ribosome, permitting another TF molecule to dock on the ribosome. Partial folding of the protein segments between TF molecules may protect them from degradation.

binding at the primary site, the ribosome, and for remaining bound to the growing protein until it detaches from the ribosome.

Although there is no dispute that trigger factor protects nascent chains, there is some controversy concerning its mode of action and its possible additional tasks. One suggestion is that, as well as preventing misfolding, trigger factor functions in a static manner by actively folding proteins within a confined void that is created when it binds to the ribosome⁸. This void has never been observed experimentally. Rather, its existence was proposed by assuming that trigger factor maintains its unbound conformation when bound to the ribosome, based on a study of a chimaeric complex of the large ribosomal subunit from *Haloarcula marismortui*³ (an archaeon, which does not possess trigger factor) and the partially resolved trigger-factor binding domain (TFa) of the eubacterium *Escherichia coli*.

The mechanistic view presented by Kaiser *et al.*² challenges this idea. Kaiser *et al.* argue that binding to the ribosome stabilizes trigger factor in an activated form, which differs from its unbound conformation. Their view agrees with the TFa conformation seen in crystal structures of TFa in complex with the large ribosomal subunit when both are from the same eubacterium, *Deinococcus radiodurans*^{4,5}. These structures indicate that TFa undergoes a conformational rearrangement on binding to the ribosome, thereby exposing a sizeable hydrophobic region facing the opening of the ribosomal exit tunnel. So, binding of trigger factor to the ribosome induces the factor to take on the activated conformation necessary for its chaperone task, and the nascent chains are prevented from aggregating by attaching to the competing hydrophobic environment on the chaperone. In this way, the competing set of interactions offered by trigger factor leads to a competition with a positive outcome: elimination of misfolding or aggregation of partially folded, newly born proteins.

The finding that a hydrophobic patch becomes exposed on trigger factor only when it binds to the ribosome, and is buried inside the free trigger factor, is consistent with the low affinity of the free factor for isolated, fully folded, hydrophobic proteins, compared with the high affinity of the ribosome-bound factor for the hydrophobic patches of growing nascent chains.

Encapsulating the emerging nascent chain within a void created by the bound trigger factor, as suggested for the modelled structure of the chimaeric complex³, presents an additional inconsistency with Kaiser and colleagues' dynamic scheme. The rather strong binding implied by the formation of such a void would hamper the smooth dissociation of the complex of trigger factor and the nascent chain from the ribosome, which is an important element in Kaiser and colleagues' proposed dynamic mechanism. However, in a partially enclosed protective space, as observed in the

eubacterial complex^{4,5}, trigger factor should be loosely bound to the ribosome and therefore can readily detach from it.

How universal is the action of trigger factor? How, for instance, does it compare with the situation in yeast? Although neither of yeast's ribosome-associated chaperones shares any amino-acid sequence similarity with the eubacterial trigger factors, yeast ribosomes can recruit trigger factor through an interaction with the yeast relative of L23, the bacterial ribosomal protein that interacts with trigger factor⁹. Does this finding indicate a common mechanism? As yet, there is no definitive answer to this. Perhaps the fact that trigger factor is exclusive to eubacteria relates to a unique property of the eubacterial protein L23, namely, the extended loop that penetrates deep into the ribosome exit tunnel and exposes a sticky hydrophobic patch on its wall. The tip of this loop might undergo conformational rearrangements when TFa binds to the ribosome¹⁰. Hence, in eubacteria, protein L23 seems to have a key role not only in trigger-factor binding, but also in the dynamic control of the nascent protein.

Further questions include whether trigger factor has additional functions, and what the conformational state of the emerging nascent chain is when it enters trigger factor's shelter, as it may already be partially folded¹¹. If it is, does this mean that ribosomes possess inherent chaperone activity themselves? Partial answers are available¹², but much work is required before we have the full picture.

Ada Yonath is in the Department of Structural Biology, Weizmann Institute, Rehovot 76100, Israel. e-mail: ada.yonath@weizmann.ac.il

1. Hartl, F. U. & Hayer-Hartl, M. *Science* **295**, 1852–1858 (2002).
2. Kaiser, C. M. *et al.* *Nature* **444**, 455–460 (2006).
3. Ferbitz, L. *et al.* *Nature* **431**, 590–596 (2004).
4. Baram, D. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 12017–12022 (2005).
5. Schlutzen, F. *et al.* *Structure (Camb.)* **13**, 1685–1694 (2005).
6. Patzelt, H. *et al.* *Biol. Chem.* **383**, 1611–1619 (2002).
7. Blaha, G. *et al.* *J. Mol. Biol.* **326**, 887–897 (2003).
8. Hoffmann, A. *et al.* *J. Biol. Chem.* **281**, 6539–6545 (2006).
9. Rauch, T. *et al.* *Mol. Microbiol.* **57**, 357–365 (2005).
10. Baram, D. & Yonath, A. *FEBS Lett.* **579**, 948–954 (2005).
11. Woolhead, C. A., Johnson, A. E. & Bernstein, H. D. *Mol. Cell* **22**, 587–598 (2006).
12. Amit, M. *et al.* *FEBS Lett.* **579**, 3207–3213 (2005).

Correction

In the News & Views article "Climate change: The north-south connection" by Eric J. Steig (*Nature* **444**, 152–153; 2006), there was an erroneous temperature inversion. The correct statement is that "there is a strong linear relationship between the magnitude of warming in Antarctica and the duration of the cold period that precedes each abrupt warming event in Greenland." And in the following passage, "The authors' explanation is simple: the duration of the warm periods in Greenland reflects the duration of reduced MOC, and hence the amount of heat retained in the Southern Ocean", "warm" should read "cold".

BRIEF COMMUNICATIONS

Mullite and the mystery of Hessian wares

Crucibles popular in the Middle Ages owed their success to an ingredient used in modern ceramics.

Crucibles from the German region of Hesse have been used since the late Middle Ages by alchemists, chemists, assayers, minters and metallurgists^{1–3}, but the factors responsible for their superior quality are unknown and several historically documented attempts to replicate their construction have failed^{2,4,5}. Here we show that the secret behind the remarkable properties of these early crucibles is mullite, an aluminium silicate that is now widely used in modern advanced ceramics.

Hessian crucibles (Fig. 1a) were established by the fifteenth century and dominated the international market, reaching Scandinavia, Britain, Portugal and even Jamestown in the American colony of Virginia^{1–3,6}. We analysed 50 Hessian and non-Hessian crucibles from 10 archaeological sites in Europe and America in an attempt to explain the superior properties of the Hessian vessels, or, as Robert Plot put it in 1677, “the mystery of the Hessian wares”⁴.

Petrographic and chemical analyses revealed that the Hessian vessels were made to a standardized recipe, in which a very lean kaolinitic clay was used, tempered with almost pure quartz sand, then shaped on a potter’s wheel and fired to very high temperatures^{3,6}.

The ceramic matrices of these crucibles have an exceptionally high alumina content (mean, 36.9 wt% $\pm$ 0.39 s.d.), with the sum of the alkali and earth-alkali oxides being just about 2 wt% (see supplementary information). This would render the vessels extremely resistant to high temperatures. In addition, Hessian fabrics contain 20–40 vol% of subangular or spheroidal quartz grains, resulting from the mixing of clay with sand. Non-plastic inclusions, notably quartz, can significantly increase the toughness and thermal-shock resistance of ceramics⁷. As a result, a Hessian crucible would be less prone to cracking if struck while its contents were being stirred, or immediately after its removal from a hot furnace.

However, these factors are not exclusive to Hessian wares. We believe that a key stage in the manufacture of Hessian crucibles, which sets them apart from their competitors, was high-temperature firing in the potter’s kiln. Scanning electron microscopy and X-ray diffraction of unused Hessian fabrics reveal complete matrix vitrification and the presence of mullite and tridymite, which most probably crystallized from the decomposition of kaolinite at temperatures between 1,100 °C and 1,200 °C (refs 8,9).

These temperatures were unusual in Europe

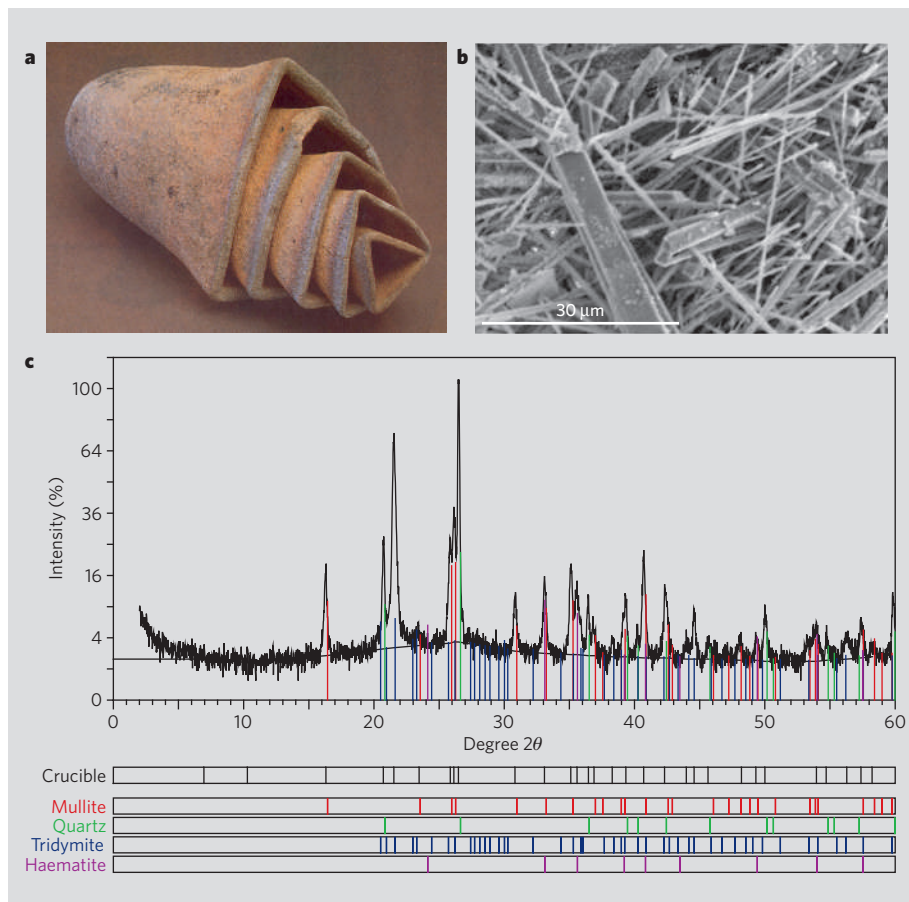


Figure 1 | Microstructure and composition of Hessian crucibles. **a**, Nest of triangular crucibles recovered in Hesse. (Photograph by S. Häpe, courtesy of H.-G. Stephan.) **b**, Electron micrograph of a cross-section through an unused Hessian crucible (original magnification, $\times 2,000$). The sample was etched with 12% hydrofluoric acid for 25 min. The image shows a fine network of mullite needles, which are very strong and resistant to temperature and corrosion. The reputation for high performance held by the Hessian wares is attributable to this mullite. (Image obtained with assistance from P. Connolly and K. Reeves.) **c**, X-ray diffraction spectrum of another unused Hessian vessel, also showing the presence of mullite (red), as well as tridymite (blue), quartz (green) and haematite (purple).

for firing pottery, with the exception of Rhinish salt-glazed stoneware, a product of the same region¹⁰. A long, high-temperature pre-firing would serve as a thermal-stability test for the crucible before delivery to the user. More important, this firing led to the formation of synthetic mullite (Fig. 1b,c), which we believe was the crucial secret behind the quality of Hessian crucibles.

Mullite ($\text{Al}_6\text{Si}_2\text{O}_{13}$) has been developed for a wide range of applications in modern ceramics, including for building and optical materials, and for thermal-protection systems and in liners for aircraft engines. Some properties of

mullite relevant to its ancestral application are low thermal expansion (and a corresponding impressive thermal-shock resistance), high creep resistance, strength at high temperatures, and an outstanding stability in aggressive chemical environments¹¹.

The presence of interlocking acicular mullite in the ceramic matrix of Hessian crucibles would have conferred properties ideal for withstanding a range of thermal, mechanical and chemical stresses. It might even be argued that the scientific developments that led to the discovery of the elements and characterization of their thermochemical behaviour were only

possible because these standardized mullite crucibles were available to yield reliable and reproducible results.

Although unaware of the presence of this aluminium silicate in their Hessian crucibles, the producers evidently coined a very successful recipe — which explains why it was not modified, or publicized, for centuries.

Marcos Martínón-Torres*, **Thilo Rehren***,
Ian C. Freestone†

*Institute of Archaeology, University College London, London WC1H 0PY, UK
e-mail: m.martinon-torres@ucl.ac.uk

†School of History and Archaeology, Cardiff University, Cardiff CF10 3EU, UK

- Stephan, H.-G. *Großalmerode: Ein europäisches Zentrum der Herstellung von technischer Keramik II* (Glas- und Keramikmuseum Großalmerode, Großalmerode, Germany, 1995).
- Cotter, J. P. in *Everyday and Exotic Pottery from Europe c. 650–1900. Studies in Honour of John G. Hurst* (eds Gaimster, D. & Redknap, M.) 256–272 (Oxbow, Oxford, 1992).
- Martínón-Torres, M. & Rehren, Th. in *Proc. 34th Int. Symp. Archaeometry, Zaragoza, 3–7 May 2004* (ed. Pérez-Arantegui, J.) 515–524 (Institución Fernando el Católico, CSIC, Zaragoza, 2006); available at <http://www.dpz.es/ifc/libros/ebook2621.pdf>
- Plot, R. *The Natural History of Oxford-shire: Being an Essay toward the Natural History of England* (printed at the Theater in Oxford, and in London by S. Millers, Oxford, 1677).
- Percy, J. *Metallurgy: The Art of Extracting Metals from their Ores, Vol. 1: Introduction, Refractory Materials, and Fuel* (Murray, London, 1875).
- Martínón-Torres, M. Thesis, Univ. London (2005).
- Tite, M. S., Kilikoglou, V. & Vekinis, G. *Archaeometry* **43**, 301–324 (2001).
- Lee, S., Kim, Y. J. & Moon, H.-S. *J. Am. Ceram. Soc.* **82**, 2841–2848 (1999).
- Chen, C.-Y. & Tuan, W.-H. *J. Am. Ceram. Soc.* **85**, 1121–1126 (2002).
- Freestone, I. C. & Tite, M. S. in *German Stoneware 1200–1900: Archaeology and Cultural History* (ed. Gaimster, D.) 354–358 (British Museum, London, 1997).
- Schneider, H. & Komarinski, S. *Mullite* (Wiley, Weinheim, 2005).

Supplementary information accompanies this communication on Nature's website.

Received 17 August; accepted 3 November 2006.

Competing financial interests: declared none.

doi:10.1038/444437a

SYNAESTHESIA

The taste of words on the tip of the tongue

Synaesthesia is a rare familial condition involving a 'crossing' of the senses — for example, ordinary activities such as reading or listening to music may be perceived with different colours or tastes¹. Here we show that individuals who experience synaesthetic tastes that are elicited by words (who are known as lexical–gustatory synaesthetes) begin to taste an upcoming word before they can actually say it (that is, while it is still 'on the tip of the tongue'). Taste sensations in these synaesthetes are therefore triggered by thinking of the word's meaning, rather than by its sound or spelling. It is possible that conceptual thought may even be linked to perceptual experience in all of us.

Brain imaging² and inheritance patterns³ have established that synaesthesia is a genuine neurological phenomenon, but the psychological mechanisms that drive each manifestation are not yet fully understood. In the word–taste variant, pictures do not usually elicit a taste sensation without retrieval of the associated word⁴, and tastes are triggered by the corresponding food name (for example, mince taste by the word *mince*) as well as by words that share phonemes with that food name (such as *prince*)^{4,5}.

Given this phonological influence, it might be assumed that explicit processing of phonological word-forms (from auditory input, for example) is a necessary stage for stimulating synaesthetic tastes. Alternatively, taste could be experienced directly from the word's meaning (that is, from abstract semantic information encoded in long-term memory, or its lemma⁶). In this case, the role of phonology would be restricted to the stages in development when word–taste associations are being established (that is, determining which words become associated with which tastes), rather than serving to trigger the taste itself.

We investigated this second hypothesis by testing whether synaesthetic tastes could be induced in 'tip-of-tongue' (TOT) states⁷, when only word meaning, but not the phonological word-form, is available for processing. Six cases

with this rare type of synaesthesia were studied (for methods, see supplementary information). Each demonstrated the behavioural hallmark of synaesthesia⁸ by showing a significantly higher test–retest consistency over more than one year, compared with a control group tested after only two weeks (for details, see supplementary information).

In a TOT picture-naming task, the participants were shown images of unusual objects (a platypus, for example) to induce the TOT state, in which the word required was known but there was a temporary inability to recall it; they were then questioned about the 'taste' of the target word. Out of 550 trials, 89 induced a TOT state in which the synaesthete was striving to name the experimenter's intended target word. Seventeen of these were accompanied by anticipatory tastes, 15 of which occurred in a complete TOT state — where neither the word itself nor any constituent letters could be recalled.

In all instances, the anticipated taste corresponded to the correct taste of the target word, as verified afterwards by the participant (confirmation stage) and then again more than one year later in a surprise retest (see supplementary information). For example, one participant tasted tuna fish when the word *castanets* was on the tip of her tongue: she then named tuna as the taste associated with the spoken item *castanets* in the confirmation stage, and reproduced this same association in the surprise retest 1.1 years later.

The anticipated taste associations named in the TOT state are unlikely to have been spuriously generated, because the synaesthetes experienced their correct anticipatory tastes significantly more often than would be predicted by chance (all *P* values less than 0.02), as estimated from the baseline frequencies with which each taste occurred for each synaesthete (for details, see supplementary information).

We conclude that perceptual experience, for these synaesthetes, is one component of the representation of the meaning of words.

As synaesthetic taste can be induced in TOT states in the absence of phonological information, we propose a model in which pathways exist between word lemmas and taste centres. (Phonological factors may nevertheless developmentally determine the nature of a taste — that is, the tastes associated with food names and their phonological neighbours.)

These pathways may operate in everyone, but be exceptionally active in synaesthetes: other variants of synaesthesia (tone–colour, for example) are known to rely on universal cognitive mechanisms⁹, and functional magnetic resonance imaging indicates that merely imagining a taste can activate the area of the normal brain associated with taste¹⁰. Lexical–gustatory synaesthesia may therefore represent an exaggeration of normal mechanisms that link linguistic thought and sensory perception.

Julia Simner*, **Jamie Ward†**

*Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK
e-mail: j.simner@ed.ac.uk

†Department of Psychology, University College London, London WC1E 6BT, UK

- Ward, J. & Mattingley, J. B. *Cortex* **42**, 129–136 (2006).
- Nunn, J. A. et al. *Nature Neurosci.* **5**, 371–375 (2002).
- Ward, J. & Simner, J. *Perception* **34**, 611–623 (2005).
- Ward, J. & Simner, J. *Cognition* **89**, 237–261 (2003).
- Ward, J., Simner, J. & Auyeung, V. *Cogn. Neuropsychol.* **22**, 28–41 (2005).
- Kempen, G. & Huijbers, P. *Cognition* **14**, 185–209 (1983).
- Burke, D. M., MacKay, D. G., Worthley, J. S. & Wade, E. *J. Mem. Lang.* **30**, 542–579 (1991).
- Rich, A. N., Bradshaw, J. L. & Mattingley, J. B. *Cognition* **98**, 53–84 (2005).
- Ward, J., Huckstep, B. & Tsakanikos, E. *Cortex* **42**, 264–280 (2006).
- Levy, L. M., Henkin, R. I., Lin, C. S., Finley, A. & Schellinger, D. *J. Comput. Assist. Tomogr.* **23**, 499–505 (1999).

Supplementary information accompanies this communication on Nature's website.

Received 25 August; accepted 3 November 2006.

Competing financial interests: declared none.

doi:10.1038/444438a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

ASTROCHEMISTRY

Complex organic matter in Titan's aerosols?

Arising from: G. Israël *et al.* *Nature* **438**, 796–799 (2005)

On 14 January 2005, the Huygens probe entered the atmosphere of Titan after a seven-year interplanetary flight as part of the Cassini mission to Saturn. Huygens carried, among other instruments, an aerosol collection and pyrolysis (ACP) device¹. Its designers, Israël *et al.*², now claim to have detected complex organic matter in two aerosol samples collected at different altitudes (130–35 km and 25–20 km, respectively), on the basis of their detection of ammonia (NH₃) and hydrogen cyanide (HCN) when the sample oven was heated to 600 °C. However, the authors' remarkable conclusions, which would have far-reaching consequences for our understanding of the chemical environment prevailing on Saturn's largest moon, are not supported by their limited data.

The claim by Israël *et al.*² to have detected NH₃ is based on the signal they obtain at a mass-to-charge ratio (*m/z*) of 17 in the 18 mass spectra (see their Fig. 3) recorded for sample 2 after the sample oven had been heated to 600 °C and swept with an isotope of nitrogen (¹⁵N₂). However, with a single exception (at 10:57:27), these signals can be accounted for by methane containing ¹³C (that is, ¹³CH₄), as calculation from the signal at *m/z* 16 (¹²CH₄) and the ¹²C/¹³C ratio (82.3) of Titan's atmospheric methane indicates³; sample 1 apparently produced similar results². Considering the small signals and large error bars indicated in Fig. 3 of ref. 2, it is not valid to conclude that ammonia has been detected in these experiments.

The claim by Israël *et al.*² that HCN could be detected is also unsupported by their data. The signal for *m/z* 27 in the mass spectrum for sample 1 is the same as that from the background (their Fig. 1) and, as the authors point out, the small peak in the spectrum for sample 2 (their Fig. 2) may well be a fragment of ethane or ethylene. The signals above *m/z* 27 of these known constituents of Titan's atmosphere are

obliterated by those from the large amounts of light (*m/z* 28, 29) and heavy nitrogen (*m/z* 30).

These two spectra are dominated by signals at *m/z* 16, 28 and 40, which correspond to the components of the atmosphere (CH₄, ¹⁴N₂ and ⁴⁰Ar). This may have entered the sample cavity through the leaking gaskets, which had caused the loss of all data from the "ambient" and 250 °C experiments² (see section 2 and Fig. 4 of the supplementary information accompanying ref. 2). These leaks would also explain the increase in the signals for sample 2, because it was collected at lower altitude (higher pressure). There is therefore no compelling evidence that any aerosol was collected or that anything was pyrolysed in these experiments.

What the authors mean by the "complex organic matter" they claim to have detected in Titan's atmospheric aerosols is illustrated in their supplementary Fig. 6: a remarkable, detailed structure, consisting of an aromatic and a cyclohexane ring, connected and substituted by several linear and branched aliphatic chains, bearing one amino, two imino and two nitrile groups. Even though this is designated as a "probable" structure, it is not justified on the basis of the authors' dubious evidence for the presence of NH₃ and HCN.

Compounds of this type would pyrolyse to small unsaturated aliphatic and aromatic molecules. Benzene and its homologues are easy to detect by mass spectrometry at very low levels owing to their aromaticity. They give rise to abundant molecular ions at *m/z* 78 and 78 + 14*n*, respectively, which is well within the mass range (*m/z* 2–142) of the mass spectrometer aboard the Huygens probe³. To explain their absence, Israël *et al.* suggest that the spectrometer has low sensitivity above *m/z* 50.

To justify the extrapolation from NH₃ and HCN to complex organic matter of the type shown in their supplementary Fig. 6, Israël *et al.*² rely on laboratory analogues (tholins) of Titan's aerosols. These have been produced

in various laboratories ever since data from Voyager's 1980 fly-by of Titan showed that its atmosphere consists of nitrogen and a few per cent of methane. Irradiation of such mixtures with various energy sources invariably produced some amorphous materials^{1,4–6}, but these were only crudely characterized and not even partially separated for structural analysis. In support of their assertion that such a tholin on pyrolysis produces only NH₃ and HCN (but nothing else), the authors show a gas chromatogram in their supplementary Fig. 1. However, that figure shows only the region from about 2–10 min of the 60-min chromatogram, cutting off just after the emergence of HCN at 9.35 min, beyond which many other pyrolysis products are likely to emerge^{1,5}. The claim to have detected complex organic matter in Titan's atmospheric haze is therefore further undermined.

Other data obtained from the successful Cassini–Huygens mission contradict or correct some previously held notions about Titan, its origin and its environment^{3,7,8}. The idea that the observed haze must be due to aerosol particles consisting of large organic molecules of complex structure needs to be re-evaluated. The chemical nature of the haze on Titan still remains shrouded in mystery.

K. Biemann

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
e-mail: kbiemann@mit.edu

1. Israël, G. *et al.* *Space Sci. Rev.* **104**, 435–466 (2002).
2. Israël, G. *et al.* *Nature* **438**, 796–799 (2005).
3. Niemann, H. *et al.* *Nature* **438**, 779–784 (2005).
4. Khare, B. N. *et al.* *Adv. Space Res.* **4**, 59–68 (1984).
5. Coll, P., Coscia, D., Gazeau, M.-C., Guez, L. & Raulin, F. *Orig. Life Evol. Biosph.* **28**, 195–213 (1997).
6. Somogyi, A., Oh, C.-H., Smith, M. A. & Lunine, J. I. *J. Am. Soc. Mass Spectrom.* **16**, 850–859 (2005).
7. Porco, C. C. *et al.* *Nature* **434**, 159–168 (2005).
8. Zarnecki, J. C. *et al.* *Nature* **438**, 792–795 (2005).

doi:10.1038/nature05417

ASTROCHEMISTRY

Israël *et al.* reply

Replying to: K. Biemann *Nature* **444**, doi: 10.1038/nature05417 (2006)

Biemann¹ calls into question our preliminary interpretations of our experimental results². A comparison of laboratory and flight measurements should settle the uncertainties he raises. In addition to evaluating instrumental

characteristics such as the 'piston effect', a technique we used for injecting oven-gas content for gas chromatography–mass spectrometry (GC–MS), such studies should enable a wide range of possible compositions for Titan's

aerosols to be investigated. Our aerosol collection and pyrolysis (ACP) measurements in Titan's atmosphere can then be revisited.

Biemann doubts that we were detecting NH₃ and HCN after the pyrolysis steps at 600 °C, as well as questioning our deductions. If the detection is valid, our inference is clear. The data obtained from temperature sensors in flight undoubtedly show that the oven heaters did work nominally, so any volatile material collected by ACP along with the haze particles must have been vaporized during the 250 °C

heating step, and — as indicated by the reading of the oven pressure sensors — evacuated from the ACP–GC–MS. Consequently, only the refractory part of the aerosols could have remained in the oven afterwards. The detection of NH_3 and HCN in the products formed during the pyrolysis of this unknown, but refractory, material indicates that it must comprise a mixture of carbon-, hydrogen- and nitrogen-containing species.

Biemann suggests that the laboratory analogues of Titan's aerosols were only crudely characterized and not even partially separated for structural analysis; however, this is not correct. The laboratory tholins were recovered and studied using different techniques in order to determine their chemical composition and their cracking patterns, which are the parameters required for interpreting observational data^{3–7}. He also misinterprets our comments on the specific structure of the aerosols. Our supplementary Fig. 6 showing the probable structure of Titan's aerosols (ref. 2) is intended to illustrate a representative example of what could be part of the structure of the complex organic material, and not what it actually is. This structure was constructed from the nature of the identified pyrolysates: nitrile groups to produce HCN, and amino or imino groups to produce NH_3 . This

type of illustration has previously been used⁸, based on experimental work⁹.

An important puzzle that will be addressed by laboratory simulations is why there are no fragments of high relative molecular mass in the spectra of the pyrolysate formed at 600 °C, which may be explained by their destruction during transport to the detector. We must, of course, accept the possibility that we could have been misled in our preliminary interpretations by a remarkable set of coincidences, which is why we are now completing a rigorous set of laboratory calibrations before reaching a final judgement on the results of the experiment.

G. Israël¹, C. Szopa¹, F. Raulin², M. Cabane¹, H. B. Niemann³, S. K. Atreya⁴, S. J. Bauer⁵, J.-F. Brun¹, E. Chassefière¹, P. Coll², E. Condé⁶, D. Coscia², A. Hauchecorne¹, P. Millian⁷, M. J. Nguyen², T. Owen⁸, W. Riedler⁹, R. E. Samuelson¹⁰, J.-M. Siguier⁷, M. Steller¹¹, R. Sternberg², C. Vidal-Madjar¹²

¹Service d'Aéronomie UMR 7620 CNRS, ISPL, Universities Paris 6 and Versailles-Saint Quentin, Verrières-le-Buisson 91371, France

²Laboratoire Interuniversitaire des Systèmes Atmosphériques, UMR 7583 CNRS, Universities Paris 12 and Paris 7, Créteil 94010, France

³NASA Goddard Space Flight Centre, Greenbelt, Maryland 20771, USA

⁴Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

⁵Institute for Meteorology and Geophysics, University of Graz, Graz 8010, Austria

⁶Centre National d'Etudes Spatiales (CNES), Toulouse 31401, France

⁷Office National d'Etudes et de Recherches Aérospatiales, Toulouse 31055, France

⁸Institute for Astronomy, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

⁹Space Research Institute, Graz 8010, Austria

¹⁰Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA

¹¹Institut für Weltraumforschung, Österreichische Akademie der Wissenschaften, Graz 8042, Austria

¹²Laboratoire de Recherche sur les Polymères, UMR 7581 CNRS, Thiais 94320, France

1. Biemann, K. *Nature* **444**, doi:10.1038/nature05417 (2006).
2. Israël, G. et al. *Nature* **438**, 796–799 (2005).
3. Khare, B. N. et al. *Icarus* **60**, 127–137 (1984).
4. McDonald, G. D. et al. *Icarus* **108**, 137–145 (1994).
5. Coll, P. et al. *Planet. Space Sci.* **47**, 1331–1340 (1999).
6. Ramirez, S. I. et al. *Icarus* **156**, 515–529 (2002).
7. Bernard, J. M. et al. *Icarus* **185**, 301–307 (2006).
8. Thompson, W. R., Henry, T. J., Schwartz, J. M., Khare, B. N. & Sagan, C. *Origins of Life* **19**, 475–476 (1989).
9. Lichtin, D. A. & Lin, M. C. *Chem. Phys.* **104**, 325–330 (1986).

doi:10.1038/nature05418

Corrigendum**Poisson's ratio and liquid's fragility**

Spyros N. Yannopoulos and G. P. Johari

Nature **442**, E7–E8; doi:10.1038/nature04967 (2006)

In this communication, we omitted to cite two additional papers that include metal glass data and discuss the relation between melt fragility and elastic properties of metallic glasses^{1,2}. Although ref. 1 presents some similar results (although not similar conclusions), our analysis took most metal glass data from ref. 2, which was submitted (25 June 2005; see also refs 3,4) before publication of ref. 1.

1. Battezzati, L. *Mater. Trans.* **46**, 2915–2919 (2005).
2. Johari, G. P. *Phil. Mag.* **86**, 1567–1579 (2006).
3. *Workshop on Viscous Liquids and the Glass Transition IV*, 3–5 June 2005, Soeminstationen, Denmark.
4. *5th Int. Disc. Meet. Relaxations in Complex Systems*, 7–13 July 2005, Villeneuve d'Ascq, France.

doi:10.1038/nature05460

Dynamics of seismogenic volcanic extrusion at Mount St Helens in 2004–05

Richard M. Iverson¹, Daniel Dzurisin¹, Cynthia A. Gardner¹, Terrence M. Gerlach¹, Richard G. LaHusen¹, Michael Lisowski¹, Jon J. Major¹, Stephen D. Malone², James A. Messerich³, Seth C. Moran¹, John S. Pallister¹, Anthony I. Qamar^{2,‡}, Steven P. Schilling¹ & James W. Vallance¹

The 2004–05 eruption of Mount St Helens exhibited sustained, near-equilibrium behaviour characterized by relatively steady extrusion of a solid dacite plug and nearly periodic shallow earthquakes. Here we present a diverse data set to support our hypothesis that these earthquakes resulted from stick-slip motion along the margins of the plug as it was forced incrementally upwards by ascending, solidifying, gas-poor magma. We formalize this hypothesis with a dynamical model that reveals a strong analogy between behaviour of the magma–plug system and that of a variably damped oscillator. Modelled stick-slip oscillations have properties that help constrain the balance of forces governing the earthquakes and eruption, and they imply that magma pressure never deviated much from the steady equilibrium pressure. We infer that the volcano was probably poised in a near-eruptive equilibrium state long before the onset of the 2004–05 eruption.

Silicic volcanoes are famous for capricious behaviour¹, but the dome-building eruption of Mount St Helens (MSH) that began in October 2004 and continues today has exhibited prolonged, near-equilibrium behaviour that affords unique opportunities for understanding volcano dynamics. A remarkable feature of the eruption has been the persistence of nearly steady, solid-state extrusion ($1\text{--}2\text{ m}^3\text{ s}^{-1}$) accompanied by nearly periodic earthquakes with shallow ($<1\text{ km}$) focal depths. The repetitive nature of these earthquakes implies the existence of a non-destructive seismic source, and it provides an exceptional opportunity to link seismicity and extrusion dynamics^{2,3}. Here we summarize seismic, geodetic, photogrammetric, petrologic, geochemical and thermal data collected during the 2004–05 MSH eruption, and we hypothesize a mechanistic link between solid-state extrusion and nearly periodic earthquakes. A new mathematical model formalizes our hypothesis, and reveals a strong analogy between the behaviour of MSH and that of a variably damped oscillator. We exploit this analogy to probe unobserved aspects of eruption dynamics.

Eruptive behaviour and seismicity

Although MSH erupted explosively on 18 May 1980, dome-building activity that began in 2004 is consistent with the volcano's recent geologic history⁴. Over the past $\sim 4,000$ yr, MSH has extruded rock at a mean rate of $\sim 0.2\text{ m}^3\text{ s}^{-1}$ while constructing a 26 km^3 edifice composed primarily of andesite and dacite lava flows and domes and their detritus (Supplementary Fig. 1). From 1980 to 1986, a dacite dome grew episodically in the 1980 crater, and its volume ultimately reached $74 \times 10^6\text{ m}^3$ (ref. 5). From 1987 to 2004, MSH did not erupt, although at least six phreatic explosions occurred from 1989 to 1991⁶. Recurrent seismicity at depths of 2–8 km in the late 1980s and 1990s may have been associated with magma recharge, but did not lead to eruptions⁷.

The current eruption began on 1 October 2004, when a small explosion formed a $\sim 20\text{-m}$ -diameter vent through a $\sim 150\text{-m}$ -thick glacier that had grown in the southern part of the MSH crater since

1986^{8–10}. The explosion was preceded by 8 days of increasingly intense seismicity at depths $<1\text{ km}$, but deeper seismicity did not occur then and has not occurred subsequently. By 11 October, explosions had largely ceased, seismic energy release had decreased to a rate about one-tenth that of the preceding two weeks, and extrusion of solid dacite had begun⁸.

The newly extruded dacite formed a series of spines with freshly exposed surfaces coated with fault gouge (that is, granulated rock bearing multiple generations of subparallel striations and slicken-sides). The most prominent spine emerged in winter 2005 and had a remarkably smooth, symmetrical whaleback form (Fig. 1 and Supplementary Movies 1 and 2), but most spines were partly obscured as they pushed past glacial ice and previously extruded rock.

Although spines repeatedly formed and disintegrated, extrusion rates remained roughly constant. The volumetric extrusion rate inferred from intermittent photogrammetric surveys was initially $\sim 6\text{ m}^3\text{ s}^{-1}$, but by December 2004 it had declined to a value of $1\text{--}2\text{ m}^3\text{ s}^{-1}$ that was sustained for the ensuing year (Fig. 2). (Supplementary Information describes measurement methods.) Similarly, the speed of plug emergence from the vent was nearly constant over timescales ranging from several minutes to several months, and was typically $3\text{--}6\text{ m d}^{-1}$. By 15 December 2005 the volume of the resulting new lava dome was $\sim 73 \times 10^6\text{ m}^3$, implying that the average MSH extrusion rate from 1980 to 2005 was similar to the $0.2\text{ m}^3\text{ s}^{-1}$ average for the past 4,000 yr.

Small earthquakes (coda duration magnitude, $M_D < 2$) that accompanied solid-state extrusion occurred so regularly that we dubbed them 'drumbeats' (Fig. 3). The period between successive drumbeats shifted slowly with time, but was nearly always 30–300 s (Fig. 2). Seismograms showed that drumbeat waveforms typically had impulsive, high-frequency onsets and low-frequency codas, similar to those of other 'hybrid' volcanic earthquakes^{2,3}. Accurate location of drumbeat hypocentres was hindered by the geologic and topographic complexity of the MSH crater and the small number

¹US Geological Survey, Cascades Volcano Observatory, 1300 SE Cardinal Ct. #100, Vancouver, Washington 98683, USA. ²Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA. ³US Geological Survey, Denver Federal Center, Box 25046, Lakewood, Colorado 80225, USA.

[‡]Deceased.

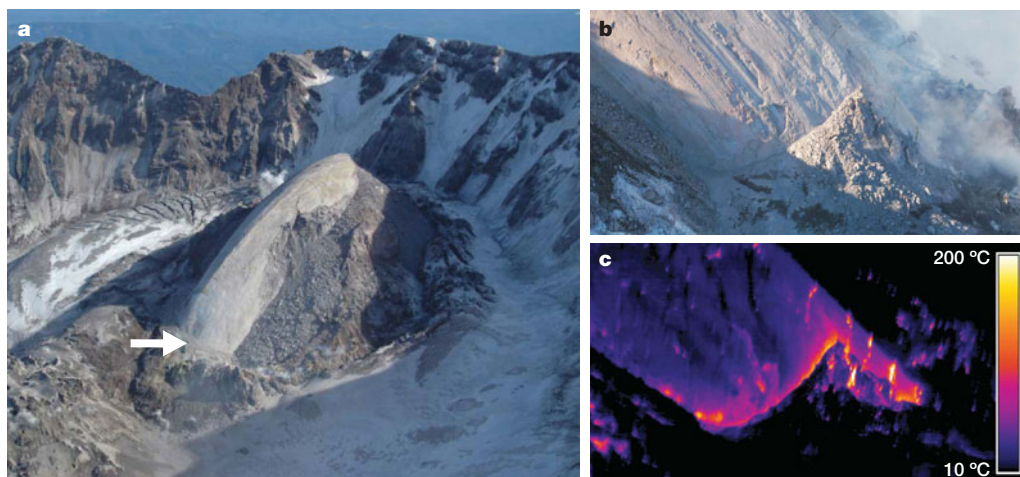


Figure 1 | Images of the whaleback spine and the surrounding MSH crater in February 2005. **a**, Oblique aerial view from the northwest. The arrow indicates the vent from which the whaleback emerged. The horizontal length

of smooth whaleback is about 380 m. **b**, Close-up view of the gouge-covered surface of the spine where it emerged from the vent, viewed from the northeast. **c**, Thermal infrared view from a perspective similar to **b**.

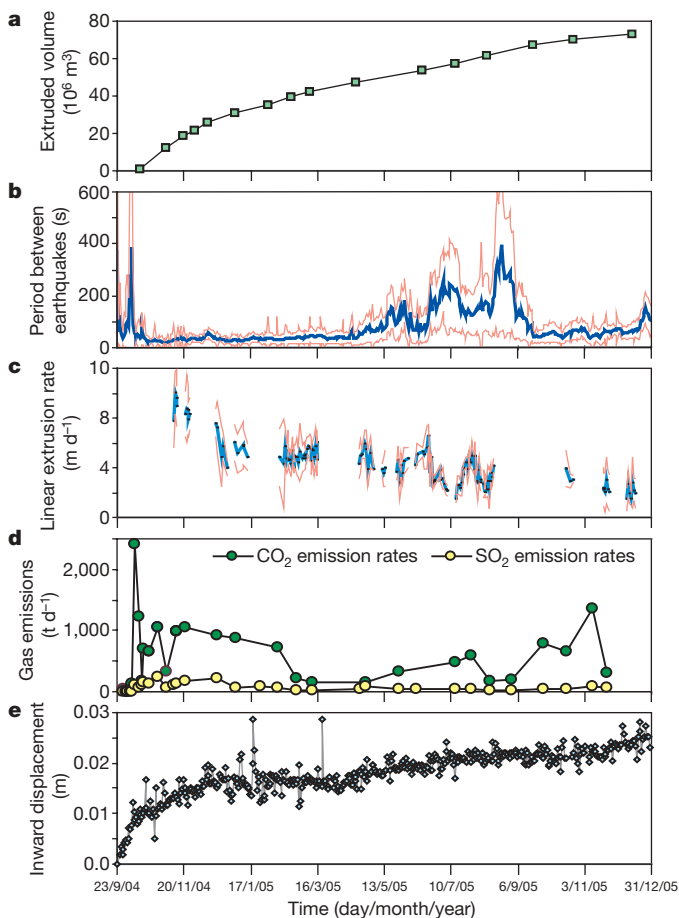


Figure 2 | Time-series data summarizing eruption behaviour.

a, Cumulative volume of extruded rock. **b**, Mean period (± 1 s.d.) between successive earthquakes recorded at station YEL (Supplementary Fig. 1), computed for 24-h sampling windows. **c**, Linear extrusion rate ± 1 s.d., inferred from repeat photography. Data gaps appear where clouds obscured the view or equipment malfunctioned. **d**, Emission rates of gaseous CO_2 and SO_2 . **e**, Cumulative inward displacement of continuous GPS station JRO1, located 9 km north of the vent (Supplementary Fig. 1). Regional displacements due to plate tectonic motion (constrained by regional GPS) have been removed from data.

of crater seismometers (Supplementary Fig. 1), but within resolution limits (~ 100 m), all drumbeats originated at depths < 1 km directly around or beneath the growing dome. Irregularly interspersed with the drumbeats were smaller and larger earthquakes ($M_D \leq 3.4$) with differing seismic signatures, but these earthquakes rarely had any effect on drumbeat occurrence.

Extruded rock, gouge and magma

Sampling of fault gouge coating the spines and of the dacite beneath the gouge was accomplished mostly by dredging from a helicopter. The gouge thickness was typically ~ 1 m and grain sizes were generally 0.001–10 mm. The gouge composition was similar to that of the newly erupted dacite, but included some constituents like those of pre-1980 MSH rocks. Airborne measurements of infrared radiation emitted by gouge emerging from the vent typically yielded temperatures of ~ 200 °C (Fig. 1), and temperatures within fissures penetrating through the gouge were no higher than 730 °C. Initial results of rheological experiments showed that even at a temperature of 1,000 °C, the new dacite deformed mostly by brittle failure rather than viscous flow, accentuating the potential for strain localization,

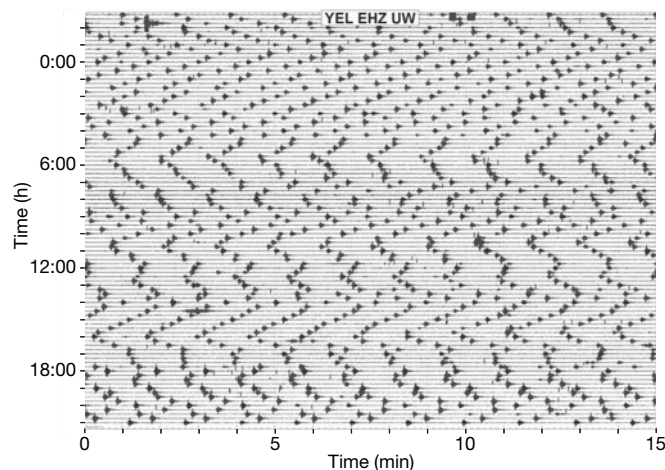


Figure 3 | Sample 24-h seismogram, illustrating regular occurrence of drumbeat earthquakes. Graph shows seismicity recorded at station YEL, located 1.5 km north of the 2004–05 vent (Supplementary Fig. 1). Time begins at 21:00 UTC on 1 December 2005, and scrolls from left to right and then top to bottom. Earthquake magnitudes were roughly 0.5–1 during this interval. (YEL/EHZ/UW is the station/component/network code for this seismic station.)

wear, and fault-gouge development^{11,12}. Room-temperature ring-shear tests on gouge specimens 7 cm thick under normal stresses of 86–191 kPa yielded steady-state (large-strain) friction coefficients μ ranging from 0.42 to 0.47, peak (small-strain) friction coefficients 1–9% larger than μ , and a roughly logarithmic decline of μ as displacement rates increased from about 10^{-6} to 10^{-4} m s⁻¹ (ref. 13).

The mineralogy and major- and trace-element compositions of the newly extruded dacite are similar to those of the youngest rocks composing the 1980s dome at MSH¹¹. Moreover, Fe-Ti oxide equilibration temperatures of samples collected in November 2004 are like those of the 1986 MSH dacites (840–850 °C). Owing to these similarities and low magmatic gas emissions, we infer that much of the 2004–05 dacite originated from magma remaining in the reservoir tapped by the 1980s eruption. Nonetheless, the long-term magma history may be complex, as extruded rocks contain phenocrysts that range widely in age and inferred source depth¹¹.

Petrologic data imply that magma solidification occurred at depths <1 km. The groundmass of the newly erupted dacite consists largely of a microlite mosaic of quartz or tridymite, sodic plagioclase and anorthoclase, with minor amounts (<15%) of high-silica rhyolite glass. The glass composition plots between the 0.1 and 50 MPa cotectics of the modified quartz–albite–orthoclase phase diagram for MSH dacites¹⁴, and the presence of tridymite further constrains the late stages of solidification to pressures of 10–20 MPa (equivalent to lithostatic pressures at depths of ~0.5–1 km)¹¹.

Volcanic gases intermittently measured by aircraft¹⁵ included CO₂, SO₂ and H₂S, and these measurements indicated the presence of magma with exsolved gas content lower than that at MSH in the early 1980s. Following the explosion of 1 October 2004, gas emission rates were <150 t d⁻¹ CO₂ and <8 t d⁻¹ H₂S, and after extrusion began on 11 October, mean emission rates increased to about 650 t d⁻¹ CO₂ and 100 t d⁻¹ SO₂ (Fig. 2). The cumulative CO₂ released through July 2005 (about 190 kt) indicates that the 2004–05 magma was gas-saturated at 8 km depth, and it constrains the gas volume fraction at that depth to <2% (assuming that the 2004–05 and 1980–86 MSH dacites have similar water contents of ~5 wt% in the rhyolitic melt fractions^{14,16,17}). Calculations using established methods¹⁶ indicate that the gas volume fraction grew during magma ascent, reached about 50% at ~1 km depth, where solidification began, and averaged ~12% between 1 and 8 km. Allowing for inevitable gas separation during extrusion, these results are consistent with observed vesicle volume fractions of 11–34% in samples of the 2004–05 dacite¹¹.

Geodetic constraints on magma source

Measured displacements of the volcano flanks and adjacent areas imply that the volume of magma evacuated from depths <10 km was considerably less than the volume of extruded rock. No evidence of systematic pre-eruption surface displacement was detected by GPS (Global Positioning System) surveys in 2000 and 2003 of a 40-station network centred on the volcano, nor by continuous operation of GPS station JRO1, located 8 km north of the volcano (Supplementary Fig. 1). Seismicity that heralded the eruption onset was accompanied by only centimetre-scale downward and inward surface displacements at JRO1 (Fig. 2). The measured displacement pattern corresponds well with predictions of an elastic half-space model¹⁸ that assumes pressure decrease within a vertically oriented, prolate spheroidal cavity with a mean depth of 8 km and volume loss of $\sim 2 \times 10^7$ m³ from 1 October 2004 to 25 November 2005. This apparent volume loss is less than one-third of the volume of rock extruded during the same period, and most of the apparent volume loss occurred before the onset of nearly steady extrusion in December 2004 (Fig. 2), implying that magma recharge from a deep (>10 km) source accompanied extrusion.

Mechanical hypothesis

To examine connections between solid-state extrusion and persistent drumbeat earthquakes at MSH, we consider mechanical processes

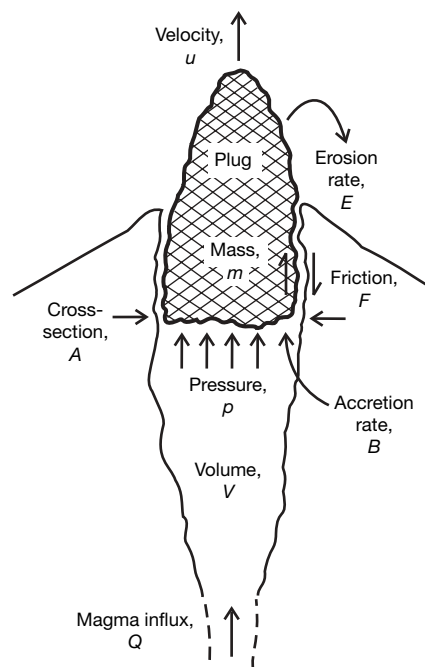


Figure 4 | Diagram of model. See text for details.

that differ from those addressed by models of erupting fluid magma^{19,20} or seismogenic motion of volcanic fluids²¹. We hypothesize that repetitive stick-slip motion along the margins of the extruding solid plug is responsible for generating drumbeat earthquakes, and that the stick-slip cycles themselves represent mechanical oscillations about equilibrium. Although prior studies have invoked stick-slip behaviour to explain cyclical behaviour during volcanic eruptions^{12,22–24}, they have not quantified the causes and character of stick-slip dynamics.

Quantification of our hypothesis requires identification of pertinent assumptions, parameters and variables (Fig. 4). We assume that magma flows into the base of an eruptive conduit (at 8 km depth, inferred from 1980–2000 seismicity) at a steady rate $Q = 2 \text{ m}^3 \text{ s}^{-1}$. An effectively rigid plug of solidified magma occupies the conduit's upper ~0.5 km (inferred from petrologic and seismic data), but the position of the plug base may change owing to upward motion and basal accretion at mass rate ρB , where ρ is magma bulk density and B is the volumetric rate of magma solidification. We assume that the plug bulk density ρ_r is constant ($2,000 \text{ kg m}^{-3}$, inferred from dome-rock specimens) and that the plug mass m can evolve owing to differences between ρB and $\rho_r E$, where E is the volumetric rate of surface erosion. For simplicity we assume that $\kappa = \rho B - \rho_r E$ is constant; thus $m = m_0 + \kappa t$, where m_0 is the initial plug mass and t is time. We estimate the horizontal cross-sectional area of the plug base, A , to be $30,000 \text{ m}^2$ (obtained by dividing $Q = 2 \text{ m}^3 \text{ s}^{-1}$ by the linear extrusion velocity $7 \times 10^{-5} \text{ m s}^{-1}$); this A implies an effective vent diameter of ~200 m, consistent with results of photogrammetric analyses (Supplementary Movie 1). We estimate the magma compressibility α_1 as 10^{-7} Pa^{-1} (appropriate for a 12% mean bubble content²⁵), and estimate the conduit wall compliance α_2 as 10^{-9} Pa^{-1} (appropriate for fractured rock). Values of α_1 and α_2 are not tightly constrained, but $\alpha_1 > \alpha_2$ is virtually certain, and we therefore infer that magma compression dominates elastic strain as the system pressurizes. Extrusion is resisted by the friction force F and the plug weight mg , where g is the acceleration due to gravity. Dependent variables that evolve during extrusion are the upward plug velocity u , magma pressure against the base of the plug p , mean magma density ρ , and volume of the magma-filled conduit V .

If drumbeat earthquakes result from incremental slip along the margins of the plug, then the slip distance per drumbeat is $\bar{u}T$, where

$\bar{u}$ is the mean extrusion velocity and T is the drumbeat period. On this basis, we infer that individual slip events typically involve plug displacements of ~ 5 mm at earthquake focal depths. Calculations test whether our mechanical hypothesis is consistent with periodic displacements of this size, although such small, brief displacements were not measurable by our instruments.

Mathematical model

We formalize our hypothesis with a model based on the following: one-dimensional conservation laws describing the evolving mass and vertical momentum of the solid plug and underlying magma; constitutive equations defining F , α_1 and α_2 ; and the assumption that Q is constant. For these conditions the governing equations reduce to:

$$\frac{du}{dt} = -g + \frac{1}{m_0 + \kappa t} (pA - \kappa u - F) \quad (1)$$

$$\frac{dp}{dt} = \frac{-1/V}{\alpha_1 + \alpha_2} (Au + RB - Q) \quad (2)$$

$$\frac{dV}{dt} = \frac{\alpha_1}{\alpha_1 + \alpha_2} (Au + RB - Q) + Q - B \quad (3)$$

where $R = 1 - (\rho/\rho_r)$ is a nearly constant coefficient determined by an isothermal equation of state, and all other variables and parameters are defined above. (Detailed derivations and discussions of these equations are provided in Supplementary Information.)

If the plug mass is constant and the basal solidification rate equals the magma influx rate, equations (1)–(3) have an equilibrium solution describing steady plug extrusion with constant magma pressure and density, conduit volume, and plug friction ($u = (Q - RB)/A$, $p = (m_0 g + F)/A$, $\rho = \rho_0$, $V = V_0$). Key dynamics questions are whether such steady states are stable, whether persistently oscillating states that can produce drumbeat earthquakes are probable, and whether initial disequilibrium states are attracted to steady or unsteady states.

Insight is gained by approximating R as constant, differentiating equation (1), combining it with equation (2), and normalizing the result to obtain an equation describing damped, forced oscillations of the scaled extrusion velocity, $u' = u/[(Q - RB)/A]$:

$$(1 + Kt') \frac{d^2 u'}{dt'^2} + 2D \frac{du'}{dt'} + \frac{u'}{V'} = \frac{1}{V'} - \frac{Kg t_0}{(Q - RB)/A} \quad (4)$$

Here $t' = t/t_0$, $V' = V/V_0$, and quantities that largely control plug motion are the characteristic time, t_0 , dimensionless mass change rate, K , and dimensionless damping, D :

$$t_0 = \frac{[m_0(\alpha_1 + \alpha_2)V_0]^{1/2}}{A}, \quad K = \frac{\kappa t_0}{m_0}, \quad (5)$$

$$D = K + \frac{1}{2} \frac{t_0}{m_0} \frac{dF}{du}$$

In the simple case in which $K = 0$, $V = V_0$, and D is constant (implying that F is a linear function of u), solutions of equation (4) indicate that u' either equals 1 (the steady equilibrium state) or oscillates about equilibrium with period $T = 2\pi t_0$. This formula yields reasonable predictions of the period between drumbeat earthquakes at MSH (Supplementary Fig. 2), but a model as simple as equation (4) with constant D cannot explain drumbeat persistence because it predicts that oscillations in extrusion rate grow unstably if $D < 0$, decay to a stable, steady state if $D > 0$, and remain unchanged only in the improbable case with $D = 0$. These inferences are modified only slightly if $K \neq 0$ or $V \neq V_0$.

Persistent oscillations are much more probable if D varies as a consequence of nonlinearly rate-dependent friction, which produces variable feedback. To illustrate effects of such feedback we use a simple nonlinear friction rule:

$$F = \text{sgn}(u) F_0 (1 - c \sinh^{-1} |u/u_{\text{ref}}|) \quad (6)$$

in which $\text{sgn}(u)$ denotes the sign of u , F_0 is the friction force at static limiting equilibrium (for details, see Supplementary Information), c is a rate-weakening parameter ($c \ll 1$), and u_{ref} is a reference value of u (Supplementary Fig. 3). This function mimics key aspects of MSH fault-gouge friction measured in experiments¹³, but it omits evolving state variables included in more elaborate friction models²⁶. Equation (6) does, however, include a fundamental type of state dependence: if oscillations lead to $u = 0$, F abruptly changes from positive to negative because gravity provides a potential for downward plug motion, which friction opposes. When F jumps, damping becomes infinite and prohibits downward settling of the plug, thereby transforming oscillatory instability to stick-slip instability (Supplementary Fig. 4).

Numerical solutions of equations (1)–(3) using a Runge-Kutta method²⁷ and incorporating equation (6) show how stick-slip cycles are regulated during extrusion. In the examples discussed here (with $K = 0$, $B = Q = 2 \text{ m}^3 \text{ s}^{-1}$, $T = 10 \text{ s}$ and $u_{\text{ref}} = 0.1(Q/A)$), values of D at the equilibrium slip rate $u_0 = Q/A$ completely determine the behaviour of solutions. As the magnitude of D increases, slip events become less frequent and more abrupt, and simultaneous evolution of u and p reveals key features of the system's dynamics (Fig. 5). When basal magma influx produces pressure exceeding the static equilibrium value, it triggers slip at a rate that depends on D . When the

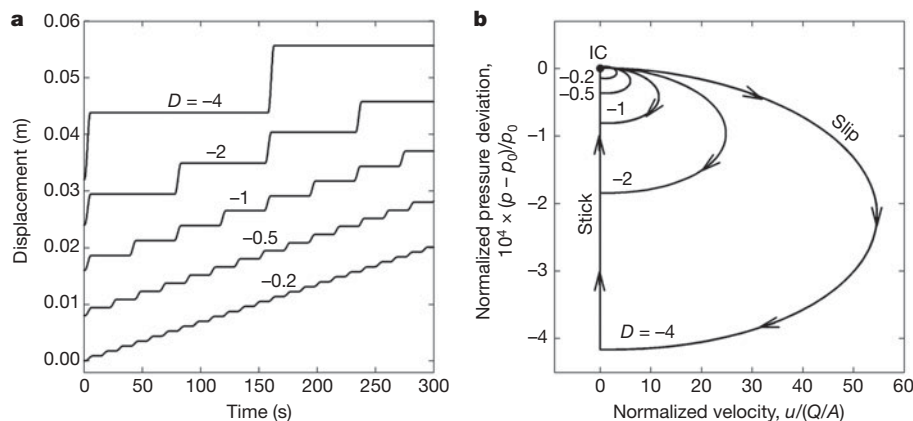


Figure 5 | Calculated stick-slip extrusion behaviour. Computations employed D values obtained by using equation (6) to find $[dF/du]_{u=u_0} = -F_0 c / u_{\text{ref}} [1 + (u_0/u_{\text{ref}})^2]^{-1/2}$ and substituting this result in equation (5). Results are depicted as time series (a), and as limit cycles in the velocity–pressure phase plane (b). Arrows in b point in the direction of

advancing time. Initial conditions (IC) for these calculations represent static limiting equilibrium specified by $u = 0$, $p = p_0 = (m_0 g + F_0)/A = 1.2936 \times 10^7 \text{ Pa}$, $V = V_0 = 6.32 \times 10^5 \text{ m}^3$ and $\rho = \rho_0 = 2,000 \text{ kg m}^{-3}$.

effects of diminishing magma pressure and plug inertia no longer exceed the effects of gravity and friction, slip terminates and stick begins. Magma pressure then rebuilds until it triggers additional slip.

Computations with $D = -2$ produce stick-slip cycles with amplitudes and periods similar to those inferred for MSH plug extrusion (Fig. 5 and Supplementary Fig. 5). With $D = -2$, individual slip events entail about 5 mm of displacement in about 5 s, maximum slip rates of $\sim 2 \text{ mm s}^{-1}$, and about 10^8 J of work done against friction. Attendant fluctuations in magma pressure are $< 0.02\%$ of p_0 , equivalent to only $\sim 2,400 \text{ Pa}$ or $< 0.2 \text{ m}$ of static magma pressure head. This result is insensitive to variations in the parameters constituting D , provided that t_0 is unchanged (Supplementary Figs 6 and 7), and multiplied by $A = 30,000 \text{ m}^2$, the pressure change also serves as a proxy for the force drop responsible for generating seismicity ($\sim 7 \times 10^7 \text{ N}$). A large fraction of this force drop occurs almost instantaneously, facilitating radiation of seismic energy (Supplementary Fig. 5).

Discussion

All silicic volcanic eruptions that follow a period of repose begin with a plugged conduit. Commonly the conduit is rapidly cleared of solid rock, and a profoundly disequibrated state exists as magma reaches the surface. In the 2004–05 eruption of MSH, the conduit remained plugged by solid rock that was pushed upward by ascending, solidifying magma. The magma–plug system reached a sustained, near-equilibrium state characterized by nearly steady extrusion of gouge-coated dacite and nearly periodic shallow earthquakes. A dynamical model of the magma–plug system produces output consistent with the hypothesis that these drumbeat earthquakes resulted from repetitive stick-slip motion along the margins of the extruding plug, and that the stick-slip cycles represent mechanical oscillations about equilibrium. Application of driving force by an unusually compliant crustal body (that is, bubbly magma) enables stick-slip events to be large enough to produce earthquakes at the shallow focal depths ($< 1 \text{ km}$) observed at MSH. If gouge displacement were driven by a stiffer body (that is, solid rock), oscillations in extrusion rate would be smaller, more frequent, and more likely to be aseismic.

Evidence that the 2004–05 eruption of MSH quickly settled into a persistent, near-equilibrium, oscillatory state implies that magma pressure never deviated much from the equilibrium pressure (Supplementary Fig. 8). Moreover, absence of deep earthquakes and minimal far-field deformation imply minimal changes in the magmatic system at depth, and we infer that the volcano was probably poised in a near-eruptive equilibrium state long before the onset of the 2004–05 eruption. In many respects, the 2004–05 eruption can be viewed as a continuation of the dome-building eruption of 1980–86, whereas the explosive eruption of 18 May 1980 differed greatly because it was triggered by a 2.5 km^3 landslide that rapidly depressurized volatile-rich magma.

Received 7 April; accepted 3 October 2006.

- Sparks, R. S. J. The causes and consequences of eruptions of andesite volcanoes: opening remarks. *Phil. Trans. R. Soc. Lond. A* **358**, 1435–1440 (2000).
- Lahr, J. C., Chouet, B. A., Stephens, C. D., Power, J. A. & Page, R. A. Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. *J. Volcanol. Geotherm. Res.* **62**, 137–151 (1994).
- Neuberg, J. Characteristics and causes of shallow seismicity in andesite volcanoes. *Phil. Trans. R. Soc. Lond. A* **358**, 1533–1546 (2000).
- Mullineaux, D. R. & Crandell, D. R. in *The 1980 Eruptions of Mount St. Helens, Washington* (eds Lipman, P. L. & Mullineaux, D. R.) 3–15 (USGS Professional Paper 1250, US Geological Survey, Washington DC, 1981).
- Swanson, D. A. & Holcomb, R. T. in *Lava Flows and Domes* (ed. Fink, J.) 3–24 (Springer, Berlin, 1990).
- Mastin, L. G. Explosive tephra emissions at Mount St. Helens, 1989–1991; the violent escape of magmatic gas following storms? *Geol. Soc. Am. Bull.* **106**, 175–185 (1994).
- Moran, S. C. Seismicity at Mount St. Helens, 1987–1992: Evidence for repressurization of an active magmatic system. *J. Geophys. Res.* **99**, 4341–4354 (1994).
- Dzurisin, D., Vallance, J. W., Gerlach, T. M., Moran, S. C. & Malone, S. D. Mount St. Helens reawakens. *Eos* **86**, 25; 29 (2005).
- Schilling, S. P., Carrara, P. E., Thompson, R. A. & Iwatsubo, E. Y. Posteruption glacier development within the crater of Mount St. Helens, Washington, USA. *Quat. Res.* **61**, 325–329 (2004).
- Walder, J. S., LaHusen, R. G., Vallance, J. W. & Schilling, S. P. Crater glaciers on active volcanoes: hydrological anomalies. *Eos* **86**, 521; 528 (2005).
- Pallister, J. S., Reagan, M. & Cashman, K. A new eruptive cycle at Mount St. Helens? *Eos* **86**, 499 (2005).
- Tuffen, H. & Dingwell, D. Fault textures in volcanic conduits: evidence for seismic trigger mechanisms during silicic eruptions. *Bull. Volcanol.* **67**, 370–387 (2005).
- Moore, P. L., Iverson, N. R. & Iverson, R. M. Frictional properties of the Mount St. Helens gouge. In *A Volcano Rekindled: The First Year of Renewed Eruption at Mount St. Helens, 2004–2005* (eds Sherrod, D. A. & Scott, W. E.) (USGS Professional Paper, US Geological Survey, submitted).
- Blundy, J. & Cashman, K. Ascent-driven crystallization of dacite magmas at Mount St. Helens, 1980–1986. *Contrib. Mineral. Petrol.* **140**, 631–650 (2001).
- McGee, K. A., Doukas, M. P. & Gerlach, T. M. Quiescent hydrogen sulfide and carbon dioxide degassing from Mount Baker, Washington. *Geophys. Res. Lett.* **28**, 4479–4482 (2001).
- Newman, S. & Lowenstern, J. A. VolatileCalc: a silicate melt– H_2O – CO_2 solution model written in Visual Basic for Application (VBA) with Microsoft® Excel. *Comput. Geosci.* **28**, 597–604 (2002).
- Rutherford, M. J. Experimental petrology applied to volcanic processes. *Eos* **74**, 49; 55 (1993).
- Yang, X., Davis, P. M. & Dieterich, J. H. Deformation from inflation of a dipping finite prolate spheroid in an elastic half space as a model for volcanic stressing. *J. Geophys. Res.* **93**, 4249–4257 (1988).
- Melnik, O. E. & Sparks, R. S. J. Nonlinear dynamics of lava dome extrusion. *Nature* **402**, 37–41 (1999).
- Mason, R. M., Starostin, A. B., Melnik, O. E. & Sparks, R. S. J. From Vulcanian explosions to sustained explosive eruptions: the role of diffusive mass transfer in conduit flow dynamics. *J. Volcanol. Geotherm. Res.* **153**, 148–165 (2006).
- Chouet, B. Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).
- Denlinger, R. P. & Hoblitt, R. P. Cyclic eruptive behavior of silicic volcanoes. *Geology* **27**, 459–462 (1999).
- Voight, B. et al. Magma flow instability and cyclic activity at Soufriere Hills Volcano, Montserrat, British West Indies. *Science* **283**, 1138–1142 (1999).
- Ozerov, A., Ispolatov, I. & Lees, J. Modeling Strombolian eruptions of Karymsky volcano, Kamchatka, Russia. *J. Volcanol. Geotherm. Res.* **122**, 265–280 (2003).
- Mastin, L. G. & Ghiorso, M. S. A Numerical Program for Steady-state Flow of Magma-gas Mixtures through Vertical Eruptive Conduits (USGS Open-File Report 00–209, US Geological Survey, Vancouver, Washington, 2000).
- Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* (Cambridge Univ. Press, Cambridge, UK, 1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to many colleagues in the USGS Volcano Hazards Program and the Pacific Northwest Seismic Network who contributed to hazard assessment and study of the ongoing eruption at MSH.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M.I. (riverson@usgs.gov), D.Z. (dzurisin@usgs.gov) or C.A.G. (cgardner@usgs.gov).

ARTICLES

Global variation in copy number in the human genome

Richard Redon¹, Shumpei Ishikawa^{2,3}, Karen R. Fitch⁴, Lars Feuk^{5,6}, George H. Perry⁷, T. Daniel Andrews¹, Heike Fiegler¹, Michael H. Shapero⁴, Andrew R. Carson^{5,6}, Wenwei Chen⁴, Eun Kyung Cho⁷, Stephanie Dallaire⁷, Jennifer L. Freeman⁷, Juan R. González⁸, Mònica Gratacòs⁸, Jing Huang⁴, Dimitrios Kalaitzopoulos¹, Daisuke Komura³, Jeffrey R. MacDonald⁵, Christian R. Marshall^{5,6}, Rui Mei⁴, Lyndal Montgomery¹, Kunihiro Nishimura², Kohji Okamura^{5,6}, Fan Shen⁴, Martin J. Somerville⁹, Joelle Tchinda⁷, Armand Valsesia¹, Cara Woodwark¹, Fengtang Yang¹, Junjun Zhang⁵, Tatiana Zerjal¹, Jane Zhang⁴, Lluís Armengol⁸, Donald F. Conrad¹⁰, Xavier Estivill^{8,11}, Chris Tyler-Smith¹, Nigel P. Carter¹, Hiroyuki Aburatani^{2,12}, Charles Lee^{7,13}, Keith W. Jones⁴, Stephen W. Scherer^{5,6} & Matthew E. Hurles¹

Copy number variation (CNV) of DNA sequences is functionally significant but has yet to be fully ascertained. We have constructed a first-generation CNV map of the human genome through the study of 270 individuals from four populations with ancestry in Europe, Africa or Asia (the HapMap collection). DNA from these individuals was screened for CNV using two complementary technologies: single-nucleotide polymorphism (SNP) genotyping arrays, and clone-based comparative genomic hybridization. A total of 1,447 copy number variable regions (CNVRs), which can encompass overlapping or adjacent gains or losses, covering 360 megabases (12% of the genome) were identified in these populations. These CNVRs contained hundreds of genes, disease loci, functional elements and segmental duplications. Notably, the CNVRs encompassed more nucleotide content per genome than SNPs, underscoring the importance of CNV in genetic diversity and evolution. The data obtained delineate linkage disequilibrium patterns for many CNVs, and reveal marked variation in copy number among populations. We also demonstrate the utility of this resource for genetic disease studies.

Genetic variation in the human genome takes many forms, ranging from large, microscopically visible chromosome anomalies to single-nucleotide changes. Recently, multiple studies have discovered an abundance of submicroscopic copy number variation of DNA segments ranging from kilobases (kb) to megabases (Mb) in size^{1–8}. Deletions, insertions, duplications and complex multi-site variants⁹, collectively termed copy number variations (CNVs) or copy number polymorphisms (CNP), are found in all humans¹⁰ and other mammals examined¹¹. We defined a CNV as a DNA segment that is 1 kb or larger and present at variable copy number in comparison with a reference genome¹⁰. A CNV can be simple in structure, such as tandem duplication, or may involve complex gains or losses of homologous sequences at multiple sites in the genome (Supplementary Fig. 1).

An early association of CNV with a phenotype was described 70 yr ago, with the duplication of the *Bar* gene in *Drosophila melanogaster* being shown to cause the Bar eye phenotype¹². CNVs influence gene expression, phenotypic variation and adaptation by disrupting genes and altering gene dosage^{7,13–15}, and can cause disease, as in microdeletion or microduplication disorders^{16–18}, or confer risk to complex disease traits such as HIV-1 infection and glomerulonephritis^{19,20}. CNVs often represent an appreciable minority of causative alleles

at genes at which other types of mutation are strongly associated with specific diseases: CHARGE syndrome²¹ and Parkinson's and Alzheimer's disease^{22,23}. Furthermore, CNVs can influence gene expression indirectly through position effects, predispose to deleterious genetic changes, or provide substrates for chromosomal change in evolution^{10,11,17,24}.

In this study, we investigated genome-wide characteristics of CNV in four populations with different ancestry, and classified CNVs into different types according to their complexity and whether copies have been gained or lost (Supplementary Fig. 1). To maximize the utility of these data and the potential for integration of CNVs with SNPs for genetic studies, we performed experiments with the International HapMap DNA and cell-line collection²⁵ derived from apparently healthy individuals. The result is the first comprehensive map of copy number variation in the human genome, which provides an important resource for studies of genome structure and human disease.

Two platforms for assessing genome-wide CNV

The HapMap collection comprises four populations: 30 parent–offspring trios of the Yoruba from Nigeria (YRI), 30 parent–offspring trios of European descent from Utah, USA (CEU), 45 unrelated

¹The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²Genome Science, and ³Dependable and High Performance Computing, Research Center for Advanced Science and Technology, University of Tokyo, 4-6-1 Komaba Meguro, Tokyo 153-8904, Japan. ⁴Affymetrix, Inc., Santa Clara, California 95051, USA. ⁵The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, MaRS Centre–East Tower, 101 College Street, Room 14-701, Toronto, Ontario M5G 1L7, Canada. ⁶Department of Molecular and Medical Genetics, Faculty of Medicine, University of Toronto M5S 1A8, Canada. ⁷Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁸Genes and Disease Program, Center for Genomic Regulation, Charles Darwin s/n, Barcelona Biomedical Research Park, 08003 Barcelona, Catalonia, Spain. ⁹Departments of Medical Genetics and Pediatrics, University of Alberta, Edmonton, Alberta T6G 2H7, Canada. ¹⁰Department of Human Genetics, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA. ¹¹Pompeu Fabra University, Charles Darwin s/n, and National Genotyping Centre (CeGen), Passeig Marítim 37-49, Barcelona Biomedical Research Park, 08003 Barcelona, Catalonia, Spain. ¹²Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. ¹³Harvard Medical School, Boston, Massachusetts 02115, USA.

Japanese from Tokyo, Japan (JPT) and 45 unrelated Han Chinese from Beijing, China (CHB). Genomic DNA from Epstein–Barr-virus-transformed lymphoblastoid cell-lines was used.

Two technology platforms were used to assess CNV (Fig. 1): (1) comparative analysis of hybridization intensities on Affymetrix GeneChip Human Mapping 500K early access arrays (500K EA), in which 474,642 SNPs were analysed; and (2) comparative genomic hybridization with a Whole Genome TilePath (WGTP) array that comprises 26,574 large-insert clones representing 93.7% of the euchromatic portion of the human genome²⁶.

Stringent quality control criteria were set for each platform and experiments were repeated for 82 individuals on the WGTP and 15 individuals on the 500K EA platforms. The quality of the final data sets was assessed by the standard deviation among $\log_2$ ratios of autosomal probes (after normalization and filtering for cell-line artefacts), which for the WGTP platform was 0.047 (Supplementary Fig. 2) and for the 500K EA platform was 0.220, both of which are improvements on published data^{8,27}.

The different nature of the two data sets required the development of distinct algorithms to identify CNVs. In essence, these algorithms segment a continuous distribution of intensity ratios into discrete regions of CNV. To train the threshold parameters, we attempted to validate experimentally 203 CNVs that had been defined with varying degrees of confidence in two well-characterized genomes^{4,5,7} (NA10851 and NA15510). By performing technical replicate experiments on both platforms we assessed the proportion of CNV calls that were false positives for different algorithm parameters across a

set of experiments representing the spectrum of data quality. The threshold parameters for both algorithms were set to achieve an average false-positive rate per experiment beneath 5% (Methods; see also Supplementary Methods, Supplementary Tables 1–4 and refs 26, 28).

Because all DNAs were derived from lymphoblastoid cell lines, we differentiated somatic artefacts (such as culture-induced rearrangements and aneuploidies) from germline CNVs. We karyotyped all available 268 HapMap cell lines (Supplementary Table 5) and sought evidence for chromosomal abnormalities in the WGTP and 500K EA intensity data. We identified 30 cell lines with unusual chromosomal constitutions (Supplementary Table 5 and Supplementary Fig. 3), and removed the aberrant chromosomes from further analyses. Chromosomes 9, 12 and X seemed to be particularly prone to trisomy. For a cell line with mosaic trisomy of chromosome 12, we confirmed by array comparative genomic hybridization that this trisomy was not apparent in blood DNA from the same individual (Supplementary Fig. 4). Furthermore, we sought signals of somatic deletions within the SNP genotypes of HapMap trios. A somatic deletion in a parental genome manifests as a cluster of SNPs at which alleles present in the offspring are not found in either parent⁵. We assessed all of our preliminary CNV calls in 120 trio parents and found that 17 (of 4,758) fell in genomic regions that harbour highly significant clusters of HapMap Phase II SNP genotypes compatible with a somatic deletion in a parental genome (Supplementary Table 5A, Supplementary Fig. 5 and Supplementary Note). These putative cell-line artefacts were removed from further analyses. Extrapolating

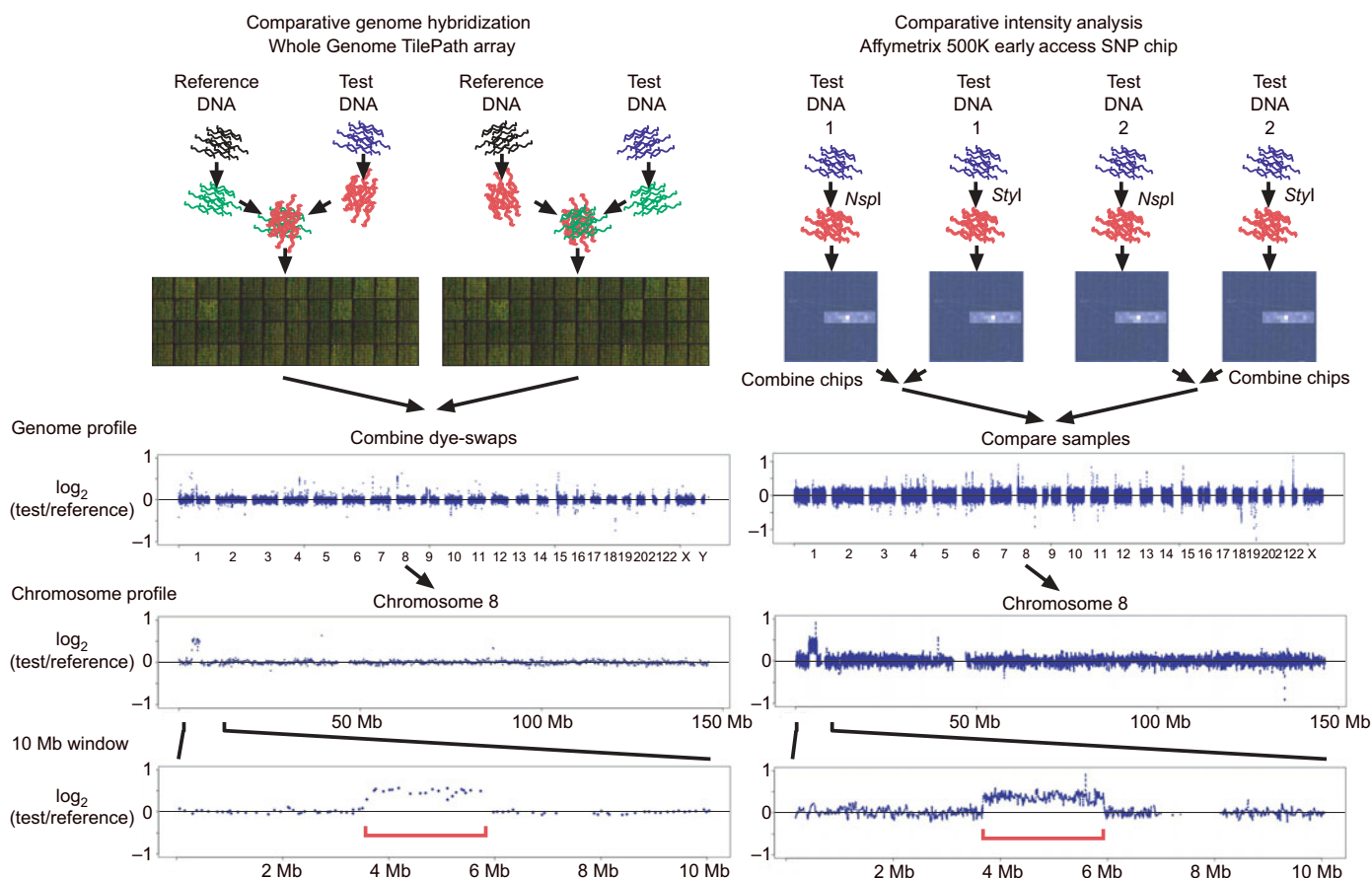


Figure 1 | Protocol outline for two CNV detection platforms. The experimental procedures for comparative genome hybridization on the WGTP array and comparative intensity analysis on the 500K EA platform are shown schematically (see Supplementary Methods for details), for a comparison of two male genomes (NA10851 and NA19007). The genome

profile shows the $\log_2$ ratio of copy number in these two genomes chromosome-by-chromosome. The 500K EA data are smoothed over a five-probe window. Below the genome profiles are expanded plots of chromosome 8, and a 10-Mb window containing a large duplication in NA19007 identified on both platforms (indicated by the red bracket).

this analysis to the entire HapMap collection suggests that less than 0.5% of the deletions we observed were likely to have been somatic artefacts.

The quality of resultant CNV calls was assessed in additional ways^{26,28}. Technical replicate experiments (triplicates for ten individuals) demonstrated that CNV calls are highly replicable (Supplementary Table 6), and that noisier experiments are characterized by higher false-negative rates, rather than higher false-positive rates (Supplementary Fig. 2). Heritability of CNVs within trios was investigated at 67 biallelic CNVs at which CNV genotypes could be inferred (Fig. 2; see also Supplementary Table 7). Of 12,060 biallelic CNV genotypes, only ~0.2% exhibited mendelian discordance, which probably reflects the genotyping error rate rather than the rate of *de novo* events at these loci. Additional locus-specific experimental validation was performed on subsets of CNVs (Supplementary Table 4). CNVs called in only a single individual (singleton CNVs) are more likely to be false positives compared with CNVs identified in several individuals. We attempted to validate 50 singleton CNVs called on only one platform (25 from each platform) and 14 singleton CNVs called on both platforms. All 14 singleton CNVs replicated by both platforms were verified as true positives, whereas 38 out of 50 of CNVs called by only one platform were similarly confirmed (false-positive rate of 24%). Extrapolating these validation rates across the entire data set suggests that only 8% (24% multiplied by the frequency of singleton CNVs called on only one platform) of the CNV regions we identify (see below) are likely to be false positives.

A genome-wide map of copy number variation

The average number of CNVs detected per experiment was 70 and 24 for the WGTP and 500K EA platforms, respectively (Supplementary Tables 8–10). Owing to the nature of the comparative analysis, each WGTP experiment detects CNVs in both test and reference genomes, whereas each 500K EA experiment detects CNV in a single genome. The median size of CNVs from the two platforms was 228 kb (WGTP) and 81 kb (500K EA), and the mean size was 341 kb and 206 kb, respectively. Consequently, the average length of the genome shown to be copy number variable in a single experiment is 24 Mb and 5 Mb on the WGTP and 500K EA platforms, respectively. The larger median size of the WGTP CNVs partially reflects inevitable overestimation of CNV boundaries on a platform comprising large-insert clones, as CNV encompassing only a fraction of a clone can be detected, but will be reported as if the whole clone was involved.

By merging overlapping CNVs identified in each individual, we delineated a minimal set of discrete copy number variable regions (CNVRs) among the 270 samples (Fig. 3; see also Supplementary Table 11). We identified 913 CNVRs on the WGTP platform and 980 CNVRs on the 500K EA platform and mapped their genomic distribution (Fig. 4). Approximately half of these CNVRs were called in more than one individual and 43% of all CNVs identified on one platform were replicated on the other. Combining the data resulted in a total of 1,447 discrete CNVRs, covering 12% (~360 Mb) of the human genome. Using locus-specific quantitative assays on a subset of regions we validated 173 (12%) of these CNVRs (Supplementary Tables 4 and 12). A minority (30%) of these 1,447 CNVRs overlapped

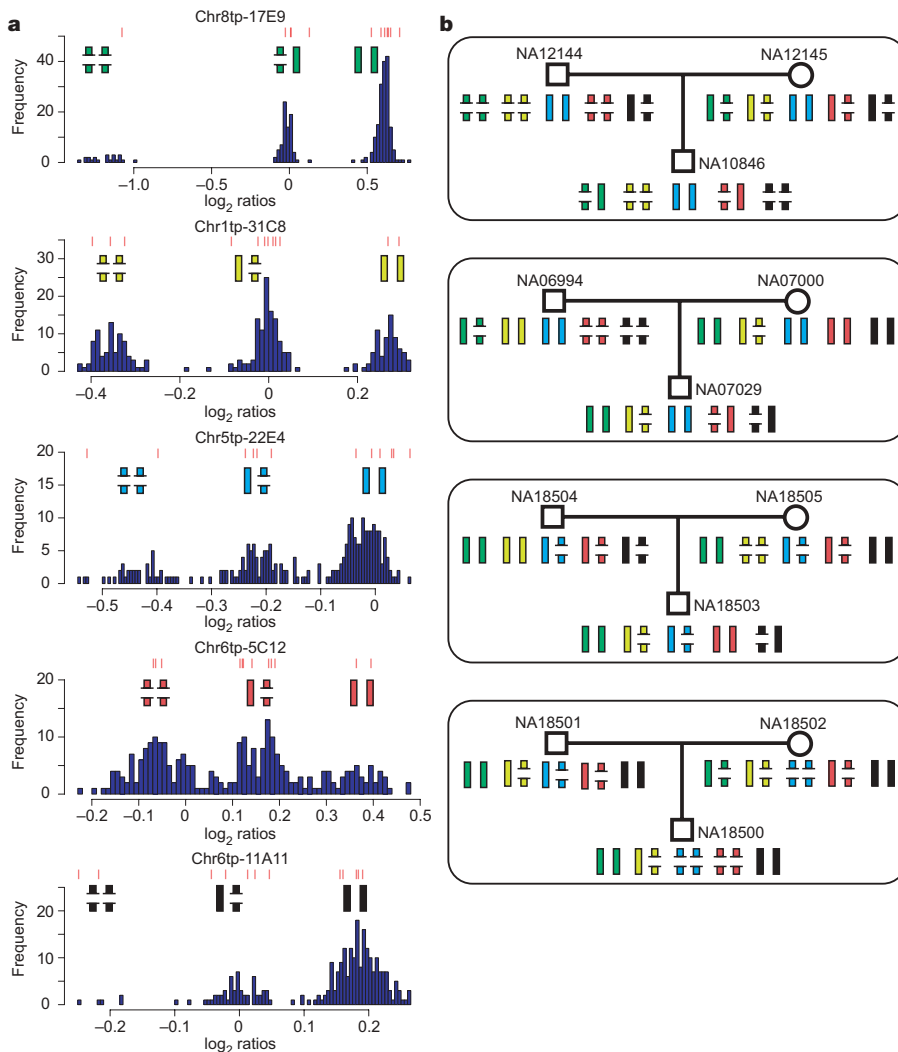


Figure 2 | Heritability of five CNVs in four HapMap trios. **a**, The distribution of WGTP $\log_2$ ratios at five CNVs with genotype information. Each histogram of $\log_2$ ratios in 270 HapMap individuals exhibits three clusters, each corresponding to a genotype of a biallelic CNV, with the two alleles depicted by broken and complete bars, representing lower and higher copy number alleles, respectively. Red lines above each histogram denote $\log_2$ ratios in the 12 individuals represented in **b**. **b**, Mendelian inheritance of five CNVs in four parent-offspring trios. The individual CNVs were genotyped from WGTP clones: green, Chr8tp-17E9; yellow, Chr1tp-31C8; blue, Chr5tp-22E4; red, Chr6tp-5C12; black, Chr6tp-11A11.

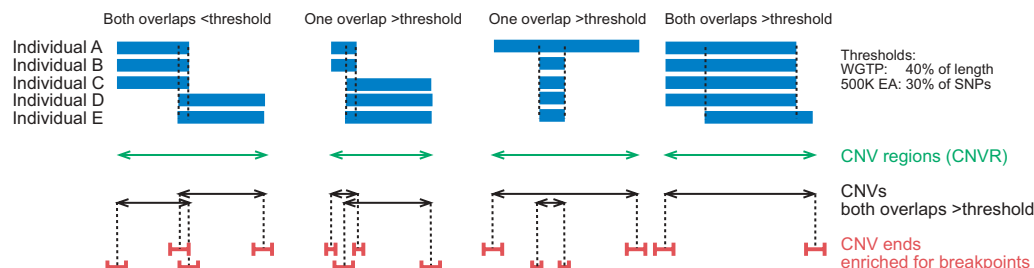


Figure 3 | Defining CNVRs, CNVs and CNV ends. Overlapping CNVs called in five individuals are shown schematically for four loci (in blue); dashed lines indicate overlap. Copy number variable regions (CNVRs) represent the union of overlapping CNVs (in green). Independent juxtaposed CNVs (in black) are identified by requiring that only individual-specific CNVs that overlap by more than a threshold proportion be merged. Intervals

encompassing CNV breakpoints (in red) are defined using platform-dependent criteria (Supplementary Methods), and contain a significant paucity of recombination hotspots^{76,77} (Supplementary Table 13), which results from the enrichment of segmental duplications within which fewer inferred recombination hotspots reside.

those identified in previous studies^{1–3,5–8,29}. Combining different classes of experimental replication revealed that 957 (66%) of the 1,447 CNVRs detected here have been replicated on both WGTP and 500K EA platforms, or with a locus-specific assay, or in another individual, or in a previous study (Supplementary Table 12). Whole-genome views of CNV show that although common, large-scale CNV is distributed in a heterogeneous manner throughout the genome (Supplementary Fig. 6), no large stretches of the genome are exempt from CNV (Fig. 4), and the proportion of any given chromosome susceptible to CNV varies from 6% to 19% (Supplementary Fig. 7).

Gaps within the reference human genome assembly have an extremely high likelihood of being associated with CNVs; out of the 345 gaps in the build 35 assembly, 48% (164 out of 345) are flanked or overlapped by CNVRs. This finding highlights the complexity in generating a reference sequence in regions of structural dynamism

and emphasizes the need for ongoing characterization of these genomic regions.

Comparing the CNVRs identified on the two platforms reveals that the WGTP and 500K EA platforms largely complement one another. The 500K EA platform is better at detecting smaller CNVs (Supplementary Fig. 8), whereas the WGTP platform has more power to detect CNVs in duplicated genomic regions (Supplementary Table 13) where 500K EA coverage is poorer³⁰.

Some CNVRs encompass two or more independent juxtaposed CNVs. For example, a small deletion found in one individual overlapping a much larger duplication in another individual was merged into a single CNVR, despite these representing distinct events. To delineate independent CNVs (CNV events) we applied more stringent merging criteria to separate juxtaposed CNVs (Fig. 3), and identified 1,116 and 1,203 CNVs on the WGTP and 500K EA

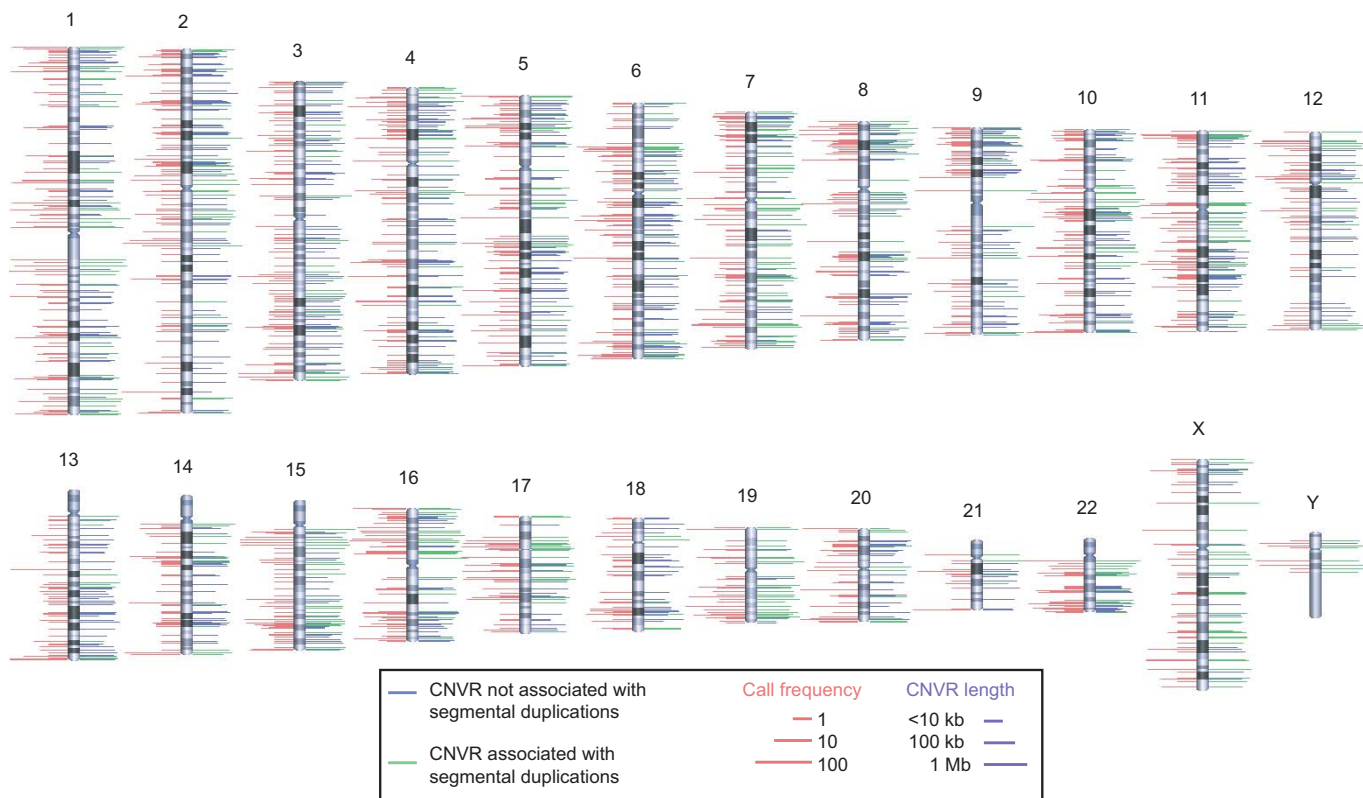


Figure 4 | Genomic distribution of CNVRs. The chromosomal locations of 1,447 CNVRs are indicated by lines to either side of ideograms. Green lines denote CNVRs associated with segmental duplications; blue lines denote CNVRs not associated with segmental duplications. The length of right-hand side lines represents the size of each CNVR. The length of left-hand side lines indicates the frequency that a CNVR is detected (minor call frequency

among 270 HapMap samples). When both platforms identify a CNVR, the maximum call frequency of the two is shown. For clarity, the dynamic range of length and frequency are log transformed (see scale bars). All data can be viewed at the Database of Genomic Variants (<http://projects.tcag.ca/variation/>).

platforms, respectively (Fig. 5; see also Supplementary Table 11). We classified these CNVs into five types: (1) deletions; (2) duplications; (3) deletions and duplications at the same locus; (4) multi-allelic loci; and (5) complex loci whose precise nature was difficult to discern. Owing to the inherently relative nature of these comparative data, it was impossible to determine unambiguously the ancestral state for most CNVs, and hence whether they are deletions or duplications. Here we adopted the convention of assuming that the minor allele is the derived allele³¹, thus deletions have a minor allele of lower copy number and duplications have a minor allele of higher copy number. Approximately equal numbers of deletions and duplications were identified on the WGTP platform, whereas deletions outnumbered duplications by approximately 2:1 on the 500K EA platform. In addition, 33 homozygous deletions (relative to the reference sequence) identified on the 500K EA platform were experimentally validated with locus-specific assays (Supplementary Table 14). Most (27 out of 33) of these have not been observed in a previous genome-wide survey of deletions⁷.

To investigate mechanisms of CNV formation, we studied the sequence context of sites of CNV. Non-allelic homologous recombination can generate rearrangements as a result of recombination between highly similar duplicated sequences^{32,33}. Segmental duplica-

tions are defined as sequences in the reference genome assembly sharing >90% sequence similarity over >1 kb with another genomic location^{34,35}. We found that 24% of the 1,447 CNVRs were associated with segmental duplications, a significant enrichment ($P < 0.05$). This association results from two factors: (1) rearrangements generated by non-allelic homologous recombination; and (2) not all annotated segmental duplications are fixed in humans, but are, in fact, CNVs. This latter point highlights the essentially arbitrary nature of defining segmental duplications on the basis of a single genome sequence (albeit derived from several individuals).

The likelihood of a CNV being associated with segmental duplications depended on its length and its classification: multi-allelic CNVs, complex CNVs and loci at which both deletions and duplications occurred were markedly enriched for segmental duplications (Fig. 5; see also Supplementary Fig. 9). This is not surprising given the role that non-allelic homologous recombination has been shown to have in generating complex structural variation³⁶, arrays of tandem duplications that vary in size³⁷, and reciprocal deletions and duplications³⁸.

The likelihood of a segmental duplication being associated with a CNV was greater for intrachromosomal duplications than for inter-chromosomal duplications, and was highly correlated with increasing sequence similarity to its duplicated copy (Supplementary Fig. 10). Non-allelic homologous recombination is known to operate mainly on intrachromosomal segmental duplications and to require 97–100% sequence similarity between duplicated copies^{33,39}.

This role for non-allelic homologous recombination in generating CNVs in duplicated regions of the genome is supported by the enrichment of segmental duplications within intervals that probably contain the breakpoints of the CNV (Fig. 3). We identified 88 CNVs from the 500K EA platform and 53 CNVs from the WGTP platform that contain a pair of segmental duplications, one at either end. These pairs of segmental duplications were biased towards high (>97%) sequence similarity, and were more frequently associated with the longest CNVs (Supplementary Fig. 11). In addition to segmental duplications, there are other types of sequence homologies that can promote non-allelic homologous recombination, for example, dispersed repetitive elements, such as *Alu* elements⁴⁰. We performed an exhaustive search for sequence homology of all kinds⁴¹ and identified 121 CNVs from the 500K EA platform and 223 on the WGTP platform that contain lengths of perfect sequence identity longer than 100 bp between either end of the CNV.

Genomic impact of CNV

Deletions are known to be biased away from genes⁵, as a result of selection. In contrast, the selective pressures on duplications are poorly understood; the existence of gene families pays testament to positive selection acting on some gene duplications over longer-term evolution⁴². We identified the different classes of functional sequence that fell within CNVRs, and tested whether they were significantly enriched or impoverished within these CNVRs compared to the entire genome (Table 1; see also Supplementary Table 13 and Supplementary Methods).

Table 1 | Functional sequences within CNVRs

Functional sequence	WGTP CNVRs	500K EA CNVRs	Merged CNVRs
RefSeq genes	2,561	1,139†	2,908†
OMIM genes	251	112†	285
Ultra-conserved elements	48†	16†	50†
Conserved non-coding elements	116,678*	55,937*	130,353*
Non-coding RNAs	57	29†	67

Statistical significance of the enrichment or paucity of functional sequences within CNVRs was assessed by randomly permuting the genomic location of autosomal CNVRs (Supplementary Methods). Significant observations are shown in bold. Note that both conserved non-coding elements⁷⁵ and CNVRs are biased away from genes, so an enrichment of conserved non-coding elements in CNVRs is not unexpected.

* Significant ($P < 0.05$) enrichment.

† Significant ($P < 0.05$) paucity.

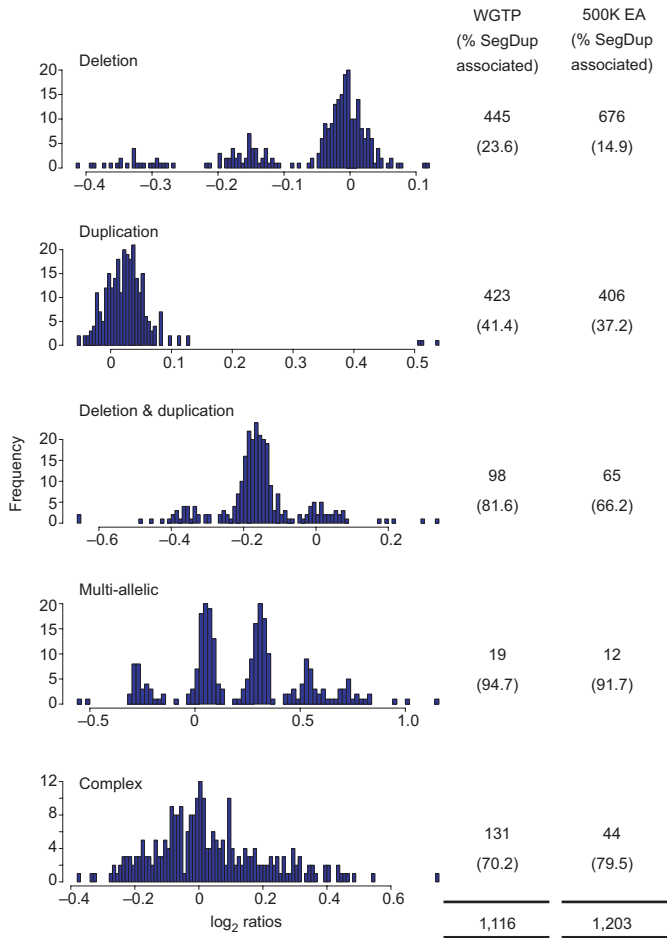


Figure 5 | Classes of CNVs. CNVs identified from WGTP and 500K EA platforms can be classified from the population distribution of log₂ ratios (exemplified with WGTP data) into five different types (see text). Biallelic CNVs (deletions and duplications) can be genotyped if the clusters representing different genotypes are sufficiently distinct. The numbers of each class of CNV identified on WGTP and 500K EA platforms are given, along with the proportion of those CNVs that overlap segmental duplications. The overall proportion of CNVRs overlapping segmental duplications was 20% and 34% on the 500K EA and WGTP platforms, respectively.

It is not possible to define precisely the breakpoints of CNVRs; therefore, some of these functional sequences might flank rather than be encompassed by CNVRs. We observed a significant paucity of all functional sequences (with the exception of conserved non-coding sequences⁴³) in CNVRs detected on the 500K EA platform, which provided the highest resolution breakpoint mapping (Table 1). Thus, CNVs are preferentially located outside of genes and ultra-conserved elements in the human genome⁴⁴. We attempted to validate experimentally 11 CNVs containing 12 ultra-conserved elements. Although all but two of the CNVs validated, only two ultra-conserved elements actually fell within these CNVs (Supplementary Table 13B), so the selection against CNV at ultra-conserved elements is likely to be even stronger than this analysis would suggest. Nevertheless, thousands of putatively functional sequences, including known disease-related genes, flank or fall within these CNVs: over half (58%) of the 1,447 CNVRs overlap known RefSeq genes, and more than 99% overlap conserved non-coding sequences⁴³.

We examined whether deletions or duplications are equally likely to encompass these different classes of functional sequences. We observed that a significantly lower proportion of deletions than duplications (identified on the 500K EA platform) overlap with the Online Mendelian Inheritance in Man (OMIM) database of disease-related genes ($P = 0.017$, chi-squared) and RefSeq genes ($P = 1.7 \times 10^{-9}$). Thus, deletions are biased away from genes with respect to duplications. The same trend was observed with ultra-conserved elements but their number is too small to provide statistical significance.

If deletions are under stronger purifying selection (which removes deleterious variants from the population) than duplications^{8,45}, then deletions should, on average, be both less frequent and smaller than duplications. Although, on average, deletions were almost threefold shorter than duplications (43 kb versus 120 kb from 500K EA), we detected no significant difference in the frequencies with which deletions and duplications were called ($P > 0.05$ using G -test for independence⁴⁶ on WGTP data). We note that our length analysis could be confounded if long duplications arise more frequently than long deletions, whereas our frequency analysis could be confounded if the power to detect duplications was lower as a result of the smaller relative change in copy number (3:2 versus 2:1).

We identified functional categories of genes that were enriched within CNVs using the Gene Ontology (GO) database (Supplementary Table 15). The most enriched GO category among genes overlapped by the 1,447 CNVRs was cell adhesion. Other highly enriched categories include sensory perception of smell and of chemical stimulus. Notably, neurophysiological processes were also a highly enriched GO category. The most highly enriched GO categories within CNVRs overlapped appreciably with those identified in a previous analysis of genes in CNVs¹⁴. Genes found in segmental duplications are known to be biased in terms of GO categories³; however, an enrichment of cell adhesion genes was also observed within CNVRs not associated with segmental duplications. We also investigated functional categories that are under-represented within CNVRs, as these might reveal classes of genes that are more likely to be dosage sensitive. We noted an impoverishment of GO categories relating to cell signalling, cell proliferation and numerous kinase- and phosphorylation-related categories. The impoverishment of these gene functions within CNVs probably reflects purifying selection acting both against the altered copy number of cell-signalling molecules vital for development and of dosage-sensitive oncogenes or tumour suppressor genes⁴⁷ that could predispose to early-onset tumorigenesis.

Copy number variation of medical relevance

In the absence of phenotypic information for HapMap donors, our data are most relevant for highlighting variable regions of the genome that warrant consideration in disease studies, rather than for immediate application to clinical diagnostics.

We found that 285 out of 1,961 (14.5%) genes in the OMIM morbid map overlapped with CNVs (Supplementary Table 16). We observed numerous examples of possible relevance to both mendelian and complex diseases. For example, the breakpoint region(s) for 12 of 25 loci involved in genomic disorders (which cause 33 different diseases) such as DiGeorge and Williams–Beuren syndromes³⁹ were found to be highly polymorphic (Supplementary Table 17). CNVs were also identified within the regions commonly deleted in DiGeorge, Smith–Magenis, Williams–Beuren, Prader–Willi and Angelman syndromes, which may be relevant for discerning uncharacterized or atypical cases. We also found CNVs at the spinal muscular atrophy and nephronophthisis loci, as expected, as these diseases are recessive in nature with relatively high carrier frequencies³³. Finally, 39 CNVs were found to reside within 500 kb of the ends of 36 chromosomal arms, which is relevant when assessing subtelomeric rearrangements in disease.

We found CNVs in genes already known to be responsible for complex traits, including *CCL3L1* and *FCGR3B*^{19,20}. Some new observations were also documented. Two CEU samples (mother and offspring) manifested a gain of CNV-95 involving the first six exons of *DISC1*, which is disrupted in schizophrenia⁴⁸. CNV-575, encompassing the LPA apolipoprotein A gene, demonstrated population variability, which may influence susceptibility to atherosclerosis. The *CRYBB2*–*CRYBB3* β -crystallin genes in CNV-1367 were observed as gains and losses in copy number in CEU and YRI samples. However, only gains were detected in Asians, leading us to speculate that variability in crystallin copy number may be linked to population differences of onset of age-related cataracts⁴⁹. Following a similar rationale, we highlight CNV-507 for possible involvement in sarcoidosis owing to its proximity to the *BTNL2* gene⁵⁰, and CNV-505 in psoriasis susceptibility, because it covers the 6p21.3 (*PSORS1*) susceptibility locus⁵¹.

We also highlight challenges in resolving genotype–phenotype correlations in complex CNVRs and how CNV detection can delineate unstable genomic regions (details in Supplementary Note including Supplementary Fig. 12). We identified patients with congenital cardiac defects and lens abnormalities⁵² that share a deleted or duplicated ~ 1 -Mb region of 1q21.1 containing pertinent candidate genes. Duplication of the same interval was observed in other cases with mental retardation²⁹. Proband can inherit the disease-associated rearrangements from unaffected parents, which underscores the variable penetrance of some diseases resulting from dosage effects^{53,54}. We found that this locus is highly duplicated, polymorphically inverted, contains assembly gaps, and is flanked by segmental duplications of variable copy number, all features being increasingly observed in CNV regions of the human genome.

Imprint of CNV on SNP genotypes

Deletions perturb patterns of marker genotypes within pedigrees⁵⁴ and these patterns highlight the location of such deletions. SNP genotype patterns characteristic of deletions are enrichment of the following: null genotypes in homozygous deletions; mendelian discrepancies in families; and Hardy–Weinberg disequilibrium within a population^{5,7}. Duplications can similarly lead to misinterpretation of marker genotypes^{53,55}, although their impact on high-density SNP maps is poorly understood.

We characterized the patterns of Phase I HapMap SNP genotypes within deletions and duplications on a chromosome-by-chromosome basis to take account of regional biases, and found that most classes of aberrant SNP genotypes were significantly enriched in both deletions and duplications on most chromosomes (Supplementary Fig. 13). We replicated the patterns of SNP genotypes within deletions (described above), and demonstrated that duplications also impact significantly on SNP genotypes. The spectrum of SNP failures enriched within duplications distinguishes them from deletions (Supplementary Fig. 13 and Supplementary Table 18). Most notably, SNPs exhibiting mendelian inconsistencies are more common within

deletions than duplications, whereas the opposite is true for SNPs in Hardy–Weinberg disequilibrium and SNPs with missing genotypes.

Cell-line artefacts also have an impact on SNP genotypes. For example, the partial deletion of 17p in NA12056 causes a distinct paucity of heterozygous SNP genotypes in that individual (one-quarter that of the population average) over several megabases of chromosome 17 (HapMap release 20).

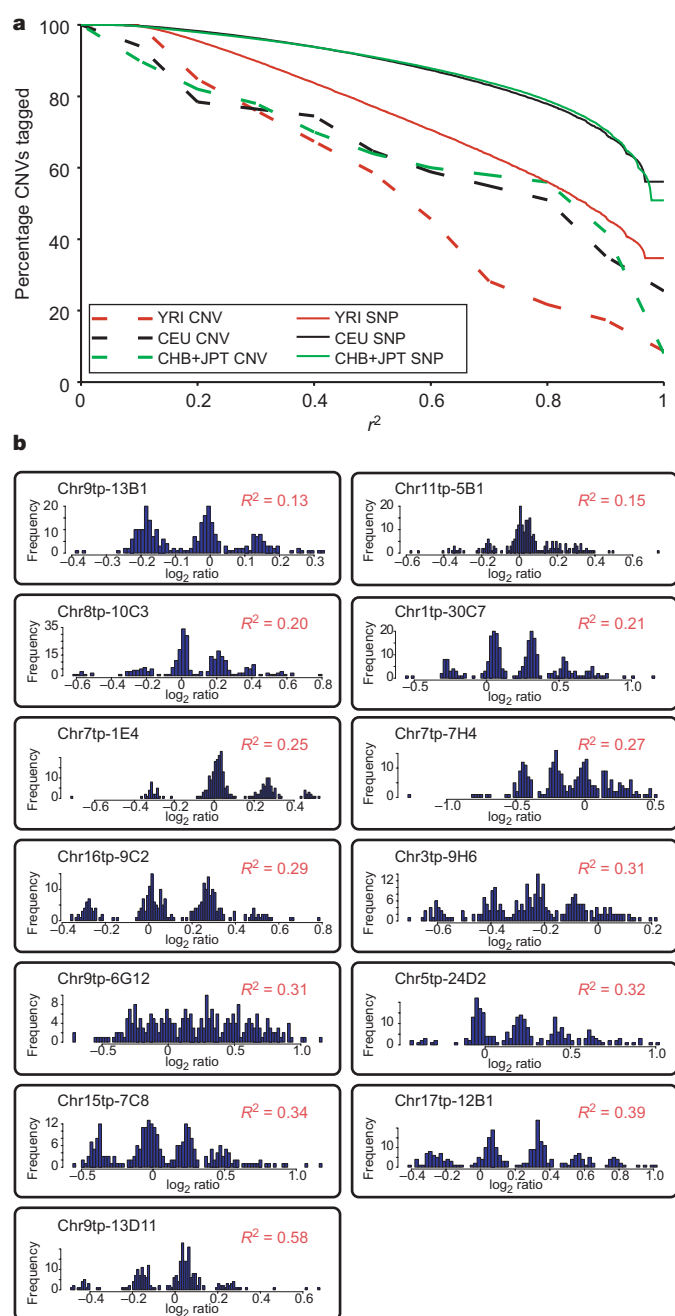


Figure 6 | Patterns of linkage disequilibrium between CNVs and SNPs. **a**, The proportion of variants that are tagged by a nearby proxy SNP (from Phase I HapMap) increases as the pairwise linkage disequilibrium (r^2) required for a proxy SNP is relaxed. This cumulative distribution is shown for both Phase I HapMap SNPs and for 65 biallelic CNVs. **b**, Histograms of the $\log_2$ ratios among all HapMap individuals are shown for 13 multi-allelic CNVs. The maximal squared Pearson correlation coefficient (R^2) observed at a neighbouring Phase I HapMap SNP—which is highly correlated with pairwise linkage disequilibrium (r^2) at biallelic CNVs (Supplementary Fig. 15)—is given for each CNV.

Linkage disequilibrium around CNVs

Indirect methods to identify causative variants, such as co-segregation of linked markers in families and genetic association with markers in linkage disequilibrium with the causative variant, are considered to be blind to the nature of the underlying mutation⁵⁶. This raises the question of whether SNP-based whole-genome association studies have the same power to detect disease-related CNVs as for disease-related SNPs. This question can be addressed by considering the maximal pairwise linkage disequilibrium (r^2) between a particular variant (CNV or SNP) and any of its neighbouring polymorphic markers. If a neighbouring marker is in high linkage disequilibrium (r^2 close to 1) with the variant of interest, that variant is ‘tagged’ by the neighbouring marker; the genotype of the variant of interest can be predicted with high probability by knowing the genotype at the ‘tagging’ marker.

Recent studies of linkage disequilibrium around CNVs have produced conflicting evidence as to the degree to which CNVs are ‘tagged’ by neighbouring SNPs^{6–8}. Here we performed a balanced comparison between the linkage disequilibrium properties of biallelic CNVs and Phase I HapMap SNPs by considering CNVs irrespective of their genomic location, and by analysing CNV genotypes (Supplementary Table 7 and Supplementary Fig. 14) of the same frequency and quality as SNP genotypes^{57,25} (see Methods for details). We quantified pairwise linkage disequilibrium around 65 biallelic CNVs using the same three analysis panels as for the Phase I HapMap (CEU, YRI and JPT plus CHB). Comparing the proportion of variants tagged by a neighbouring SNP with an arbitrary threshold of $r^2 > 0.8$ shows that whereas 75–80% of Phase I SNPs in non-African populations were tagged, only 51% of CNVs were tagged in the same populations (Fig. 6a and Supplementary Table 19). In the YRI, both SNPs and CNVs exhibited lower linkage disequilibrium, with only 22% of CNVs being tagged with $r^2 > 0.8$.

We considered three explanations for these observations of lower apparent linkage disequilibrium around CNVs than SNPs. First, some duplications might represent transposition events that would generate linkage disequilibrium around the (unknown) acceptor locus but not the donor locus. One of the genotyped CNVs is known to be a duplicative transposition⁵⁷, but evidence from *de novo* pathogenic duplications strongly suggests a preference for tandem, rather than dispersed, duplications, regardless of whether the duplication is caused by non-allelic homologous recombination⁵⁸. Second, some CNVs might undergo recurrent mutations or reversions, especially tandem duplications which are mechanistically prone to unequal crossing over, causing reversions back to a single copy¹². However, duplications were not in lower linkage disequilibrium with flanking SNPs than were deletions. Finally, we considered that CNVs might occur preferentially in genomic regions with lower densities of SNP genotypes in HapMap Phase I. We found that CNVs are enriched within segmentally duplicated regions of the genome, in which there is a paucity of genotyped SNPs owing to technical difficulties²⁵. Thus, the strongest factor decreasing apparent linkage disequilibrium around biallelic CNVs is not that linkage disequilibrium around such CNVs is necessarily lower, but that there is, on average, lower coverage of these structurally dynamic regions of the genome by SNPs genotyped in Phase I of the HapMap project.

We investigated whether the copy number of multi-allelic CNVs could be predicted reliably by nearby SNPs. We treated diploid genome copy number at multi-allelic CNVs as a quantitative trait, and asked which nearby SNPs are most predictive of this trait, and how strongly predictive these SNPs are. We identified 13 multi-allelic CNVs in which the quantitative WGTP data clearly clustered into discrete diploid genome copy numbers and quantified the predictive ability of neighbouring SNPs using the square of Pearson’s correlation coefficient (R^2) (Supplementary Fig. 15). We found that diploid copy number of multi-allelic CNVs is poorly predicted by neighbouring SNPs (Fig. 6b). It may be that combining information from

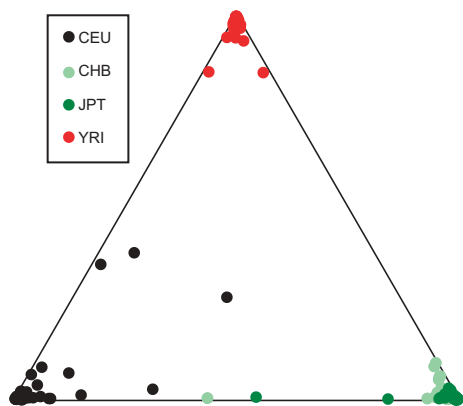


Figure 7 | Population clustering from CNV genotypes. A triangle plot showing the clustering of 210 unrelated HapMap individuals assuming three ancestral populations ($k = 3$). The proximity of an individual to each apex of the triangle indicates the proportion of that genome that is estimated to have ancestry in each of the three inferred ancestral populations. The clustering together of most individuals from the same population near a common apex indicates the clear discrimination between populations obtained through this analysis. The clustering was qualitatively similar to that obtained previously with a similar number of biallelic *Alu* insertion polymorphisms on different African, European and Asian population samples⁶⁰.

several SNPs could provide greater power for predicting diploid genome copy number at these loci.

Population genetics of copy number variation

In contrast to other classes of human genetic variation, the population genetics of copy number variation remains unexplored. The distribution of copy number variation within and among different populations is shaped by mutation, selection and demographic history. A range of polymorphisms, including SNPs²⁵, microsatellites⁵⁹ and *Alu* insertion variants⁶⁰, has been used to investigate population structure. To demonstrate the utility of copy number variation genotypes for population genetic inference we performed population clustering⁶¹ on 67 genotyped biallelic CNVs. We obtained the optimal clustering with the assumption of three ancestral populations, with the African, European and Asian populations clearly differentiated (Fig. 7). Population differentiation of individual variants is commonly estimated by the statistic F_{ST} , which varies from 0 (undifferentiated) to 1 (population-specific)⁶². The average F_{ST} for the same 67 autosomal CNVs was 0.11, very similar to that observed for all autosomal Phase I HapMap SNPs (0.13)²⁵.

Recent population-specific positive selection elevates population differentiation. To explore population differentiation at all CNVs, we devised a statistic, V_{ST} , that estimates population differentiation based on the quantitative intensity data and varies from 0 to 1, similar to F_{ST} (Supplementary Fig. 16). Estimating V_{ST} for all clones on the

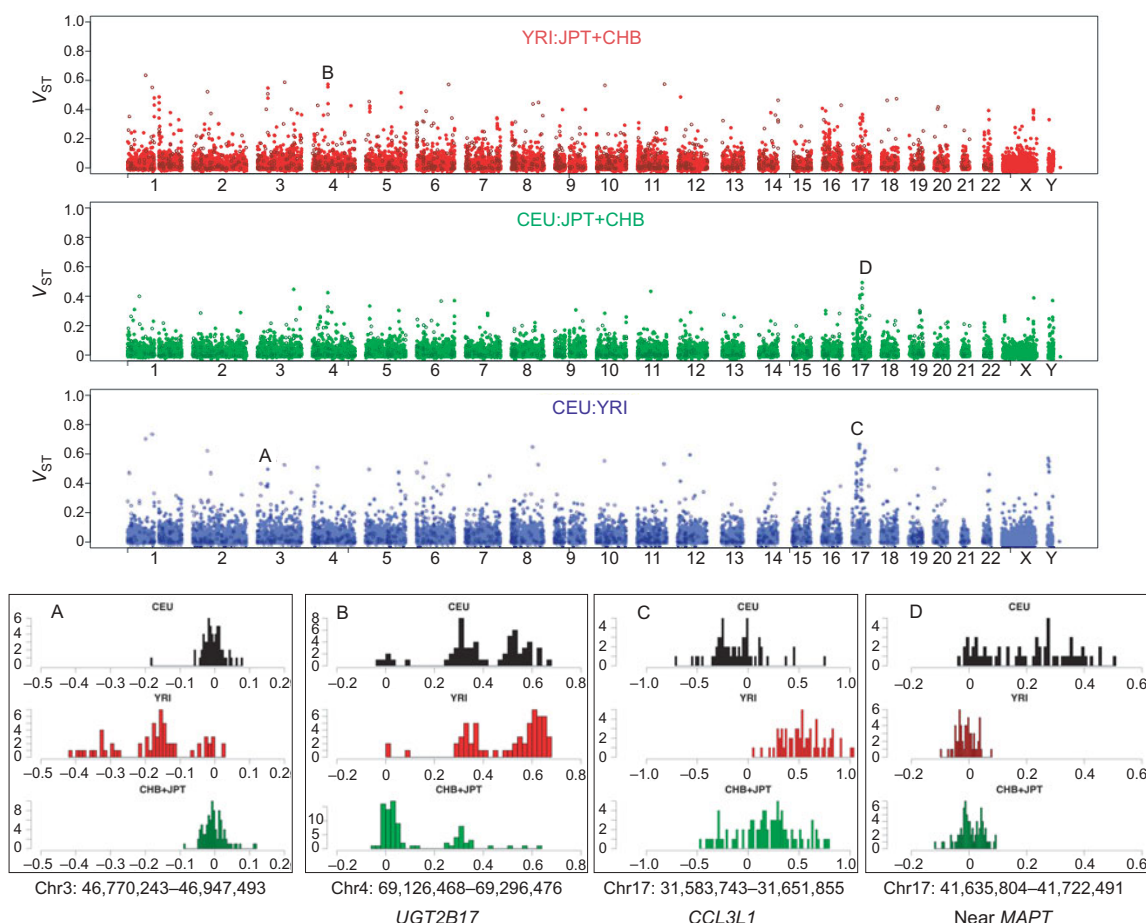


Figure 8 | Population differentiation for copy number variation. Population differentiation, estimated by V_{ST} , for each of the three population pairwise comparisons is plotted along each chromosome. For each pairwise comparison, the V_{ST} values for all clones on the WGTP platform are shown in the lighter colour with filled circles, with V_{ST} values of CNVs detected on the 500K EA platform superimposed in a darker shade with unfilled circles. Histograms showing the distributions of $\log_2$ ratios (on

the WGTP platform) among the unrelated individuals in each population are plotted for four example CNVs exhibiting high population differentiation, labelled A–D. y axis values indicate absolute frequency. Each example histogram is labelled with the chromosome coordinates of the WGTP clone, and flanking/encompassed genes are given for those CNVs mentioned in the text.

WGTP array and all CNVs on the 500K EA array revealed a number of outliers with levels of population differentiation suggestive of population-specific selective pressures (Fig. 8; see also Supplementary Table 20). Among these outliers were two CNVs previously demonstrated to have elevated population differentiation^{7,19}: *UGT2B17* is a gene encoding a UDP-glucuronosyl transferase with roles in androgen metabolism and xenobiotic conjugation^{63,64}, and *CCL3L1* is a chemokine-encoding multi-copy gene at which greater copy numbers protect against HIV-1 infection¹⁹.

Not all regions that have been under recent positive selection exhibit elevated population differentiation⁶⁵. To detect other CNVs that may have recently been under positive selection, we identified CNVRs that fell within 124 out of 752 (16%) genomic locations previously shown⁶⁶ to exhibit haplotype patterns suggestive of a partial selective sweep (Supplementary Table 21). Two of these overlapping CNVs also fell within the set exhibiting highest population differentiation, shown in Supplementary Table 20. One of these selection-associated CNVs is a duplication specific to the CEU (Fig. 8) and lies near to the *MAPT* gene, which is associated with a set of neurodegenerative disorders known as 'tauopathies'⁶⁷. Both the *MAPT* gene and the duplication lie within a chromosomal region that has recently been shown to have a complex evolutionary history, characterized by a common chromosomal inversion, deep divergence between inverted haplotypes and recent positive selection in European populations⁶⁸. We adapted methods used to identify partial selective sweeps at SNPs^{66,69} to estimate relative extended haplotype homozygosity values (REHH) on either flank of the 67 CNVs for which we extracted genotype information (Supplementary Methods). We identified no convincing signals ($P < 0.01$ on both flanks) of positive selection on any CNV in any one population, although there were weaker signals ($P < 0.05$) apparent for some CNVs (Supplementary Fig. 17 and Supplementary Table 21).

Discussion

Our map of copy number variation in the human genome demonstrates the ubiquity and complexity of this form of genomic variation. The abundance of functional sequences of all types both within and flanking areas of copy number variation suggests that the contribution of CNVs to phenotypic variation is likely to be appreciable. This prediction is underscored by the impact of copy number variation on variation in gene expression (B. Stranger and E. Dermitzakis, personal communication).

CNV assessment should now become standard in the design of all studies of the genetic basis of phenotypic variation, including disease susceptibility. Similarly important will be CNV annotation in all future genome assemblies. The identification of CNVs causing severe, sporadic diseases has been hampered by an inability to distinguish between normal and causative variants. Our CNV map, in tandem with the DECIPHER (<http://www.sanger.ac.uk/PostGenomics/decipher/>) project's sharing of copy number information on patients with rare, severe phenotypes should advance progress in this area. For mendelian genetic diseases, our data contain numerous known and candidate recessive disease alleles, and existing linkage data can be used to prioritise CNVs for further investigation.

Genetic association studies are the predominant strategy for identifying haplotypes conferring risk for complex genetic diseases. Such studies are typically based on SNP genotyping, either within candidate loci or genome-wide⁵⁶. Our analysis of linkage disequilibrium between CNVs and SNPs gives us limited optimism that CNVs influencing risk to complex disease will be detected by such approaches. The tag SNPs that we have identified (Supplementary Table 19) for specific CNVs can be used as proxies for these CNVs. Moreover, CNV-specific genotyping assays can be developed for CNVs for which tag SNPs are not readily identifiable but whose proximity to candidate genes warrants further characterization. Finally, we see great merit in mining CNV information from quantitative SNP gen-

otyping data, and in enriching future generations of genome-wide SNP genotyping platforms by targeting untaggable CNVs.

The overall utility of any map depends on its coverage and its completeness. Extrapolation based on existing data suggests that smaller deletions (<20 kb) are much more frequent than larger deletions (>20 kb)⁵, and the same may be true for duplications. Although we have generated the most complete CNV map yet described, given our lower power to detect smaller CNVs a substantial fraction of copy number variation >1 kb in size in these individuals remains to be characterized. No single available technology will capture all variation. Smaller rearrangements are amenable to detection using technologies such as sequence assembly comparisons⁷⁰, paired-end sequence relationships⁴, sequence trace analysis⁷¹ and higher-resolution tiling arrays⁵. Ultimately, it is desirable to know the precise chromosomal location and sequence content of each and every CNV. At present, generating this information requires the use of multiple experimental methods. However, in the future, the comparison of independently assembled whole-genome sequences could provide a definitive solution.

Our study of human genetic variation ties together cytogenetics, submicroscopic copy number variation and single nucleotide polymorphisms, providing a framework for future genetic studies. This framework will need to be supported by continual refinement of the reference genome sequence and robust nomenclature and databasing for structural variation—both facilitated through international collaboration—to enable further unravelling of the complexity of human genomic variation.

METHODS

CNV determination on WGTP and 500K EA arrays. The experimental methods and algorithms used for CNV calling using the WGTP and the 500K EA platforms are described in the Supplementary Methods and elsewhere^{26,28}.

500K EA and WGTP data quality assessment. In order to estimate false-positive and false-negative rates on both platforms, we used quantitative PCR to test experimentally CNVs called from replicate experiments with NA15510 (ref. 4; Supplementary Tables 1–4). With 500K EA, the average proportion of CNV calls from this sample that were false positive was 2.3% (0.33 out of 15 CNV calls), and the false-negative percentage was 24% (3.3 validated CNVs not called in any one replicate per 38 total validated CNVs)²⁸. With WGTP, we found an average false-positive proportion of 5% (3.4 out of 68.2) and a false-negative percentage of 37.8% (58 non-called out of 154 tested)²⁶. Reproducibility was also assessed by analysing ten HapMap DNAs in triplicate (Supplementary Table 6). For the 500K EA platform, on average, 80% of CNVs were called in all three replicates, 10% were called twice, and 10% were called only once. For the WGTP platform, using the replicate with the lowest number of calls as the baseline, on average 73% of CNVs were called in three experiments, 14% were called twice, and 13% were called once²⁶.

Population genetic and statistical analyses. Sixty-seven non-redundant biallelic CNVs suitable for genotyping were identified on the WGTP and 500K EA platforms. Two procedures were used to cluster intensity ratios into discreet copy number genotypes: K-means and partitioning around medoids (PAM)⁷². Pairwise linkage disequilibrium (r^2) was estimated between 65 biallelic CNVs and all filtered non-redundant Phase I HapMap SNPs within 500 kb of the CNV borders using HaploView⁷³. Population clustering was performed using STRUCTURE⁶¹ and population-specific CNVs were estimated using F_{ST} ⁶² and the new statistic V_{ST} (Supplementary Fig. 16). V_{ST} is calculated by considering $(V_T - V_S)/V_T$, where V_T is the variance in $\log_2$ ratios apparent among all unrelated individuals and V_S is the average variance within each population, weighted for population size. For REHH analysis^{69,74}, we treated each genotyped CNV as a SNP located at the CNV end, and used Phase I HapMap SNPs 500 kb upstream and downstream of the CNV using the program Sweep (<http://www.broad.mit.edu/mpg/sweep/resources.html>). See Supplementary Methods for more details.

Data release. The raw data from the 500K EA as well as the 500K commercial arrays are posted at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>), with accession numbers GSE5013 and GSE5173. WGTP data are posted at ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) with accession number E-TABM-107, and at the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/humgen/cnv/data/>). CNV calls have been released at the Database of Genomic Variants (<http://projects.tcag.ca/variation/>) integrated with all other CNV data.

Received 13 June; accepted 10 October 2006.

1. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
2. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
3. Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
4. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
5. Conrad, D. F., Andrews, T. D., Carter, N. P., Hurler, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nature Genet.* **38**, 75–81 (2006).
6. Hinds, D. A., Kloek, A. P., Jen, M., Chen, X. & Frazer, K. A. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genet.* **38**, 82–85 (2006).
7. McCarroll, S. A. *et al.* Common deletion polymorphisms in the human genome. *Nature Genet.* **38**, 86–92 (2006).
8. Locke, D. P. *et al.* Linkage disequilibrium and heritability of CNPs within duplicated regions of the human genome. *Am. J. Hum. Genet.* **79**, 275–290 (2006).
9. Fredman, D. *et al.* Complex SNP-related sequence variation in segmental genome duplications. *Nature Genet.* **36**, 861–866 (2004).
10. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7**, 85–97 (2006).
11. Freeman, J. L. *et al.* Copy number variation: new insights in genome diversity. *Genome Res.* **16**, 949–961 (2006).
12. Bridges, C. B. The Bar 'gene': a duplication. *Science* **83**, 210–211 (1936).
13. Buckland, P. R. Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann. Med.* **35**, 308–315 (2003).
14. Nguyen, D. Q., Webber, C. & Ponting, C. P. Bias of selection on human copy-number variants. *PLoS Genet.* **2**, e20 (2006).
15. Repping, S. *et al.* High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature Genet.* **38**, 463–467 (2006).
16. Inoue, K. & Lupski, J. R. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3**, 199–242 (2002).
17. Lupski, J. R. & Stankiewicz, P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet.* **1**, e49 (2005).
18. Shaw-Smith, C. *et al.* Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J. Med. Genet.* **41**, 241–248 (2004).
19. Gonzalez, E. *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
20. Aitman, T. J. *et al.* Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
21. Jongmans, M. C. *et al.* CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. *J. Med. Genet.* **43**, 306–314 (2006).
22. Singleton, A. B. *et al.* α -Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
23. Rozelet-Lecrux, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature Genet.* **38**, 24–26 (2006).
24. Feuk, L., Marshall, C. R., Wintle, R. F. & Scherer, S. W. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum. Mol. Genet.* **15** (Suppl. 1), R57–R66 (2006).
25. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
26. Fiegler, H. *et al.* Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res.* **16**, 1566–1574 (2006).
27. Ylstra, B., van den Ijssel, P., Carvalho, B., Brakenhoff, R. H. & Meijer, G. A. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res.* **34**, 445–450 (2006).
28. Komura, D. *et al.* Genome-wide detection of human copy number variations using high density DNA oligonucleotide arrays. *Genome Res.* **16**, 1575–1584 (2006).
29. de Vries, B. B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
30. Wirtenberger, M., Hemminki, K., Chen, B. & Burwinkel, B. SNP microarray analysis for genome-wide detection of crossover regions. *Hum. Genet.* **117**, 389–397 (2005).
31. Marth, G. T., Czarbarka, E., Murvai, J. & Sherry, S. T. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* **166**, 351–372 (2004).
32. Hurler, M. How homologous recombination generates a mutable genome. *Hum. Genomics* **2**, 179–186 (2005).
33. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
34. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
35. Cheung, J. *et al.* Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**, R25 (2003).
36. Aradhya, S. *et al.* Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. *Hum. Mol. Genet.* **10**, 2557–2567 (2001).
37. Neitz, M. & Neitz, J. Numbers and ratios of visual pigment genes for normal red-green color vision. *Science* **267**, 1013–1016 (1995).
38. Chance, P. F. *et al.* Two Autosomal-dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome-17. *Hum. Mol. Genet.* **3**, 223–228 (1994).
39. Lupski, J. R. & Stankiewicz, P. (eds) *Genomic Disorders: the Genomic Basis of Disease* (Humana Press, Totowa, New Jersey, 2006).
40. Deininger, P. L. & Batzer, M. A. Alu repeats and human disease. *Mol. Genet. Metab.* **67**, 183–193 (1999).
41. Kurtz, S. *et al.* REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).
42. Hurler, M. Gene duplication: the genomic trade in spare parts. *PLoS Biol.* **2**, E206 (2004).
43. Drake, J. A. *et al.* Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nature Genet.* **38**, 223–227 (2006).
44. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
45. Brewer, C., Holloway, S., Zawalski, P., Schinzel, A. & FitzPatrick, D. A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality—and tolerance of segmental aneuploidy—in humans. *Am. J. Hum. Genet.* **64**, 1702–1708 (1999).
46. Sokal, R. R. & Rohlf, J. F. *Biometry* (W. H. Freeman and Co, New York, 1994).
47. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
48. Millar, J. K. *et al.* DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* **310**, 1187–1191 (2005).
49. Sasaki, K., Sasaki, H., Jonasson, F., Kojima, M. & Cheng, H. M. Racial differences of lens transparency properties with aging and prevalence of age-related cataract applying a WHO classification system. *Ophthalmic Res.* **36**, 332–340 (2004).
50. Valentonyte, R. *et al.* Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nature Genet.* **37**, 357–364 (2005).
51. Orru, S., Giurelli, E., Carcassi, C., Casula, M. & Contu, L. Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-Kb interval around the corneodesmosin gene (CDSN). *Am. J. Hum. Genet.* **76**, 164–171 (2005).
52. Christiansen, J. *et al.* Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ. Res.* **94**, 1429–1435 (2004).
53. Lupski, J. R. *et al.* DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66**, 219–232 (1991).
54. Chance, P. F. *et al.* DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72**, 143–151 (1993).
55. Matise, T. C. *et al.* Detection of tandem duplications and implications for linkage analysis. *Am. J. Hum. Genet.* **54**, 1110–1121 (1994).
56. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).
57. Wong, Z., Royle, N. J. & Jeffreys, A. J. A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics* **7**, 222–234 (1990).
58. Woodward, K. J. *et al.* Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination. *Am. J. Hum. Genet.* **77**, 966–987 (2005).
59. Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
60. Bamshad, M. J. *et al.* Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **72**, 578–589 (2003).
61. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
62. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution Int. J. Org. Evol.* **38**, 1358–1370 (1984).
63. Turgeon, D., Carrier, J. S., Chouinard, S. & Belanger, A. Glucuronidation activity of the UGT2B17 enzyme toward xenobiotics. *Drug Metab. Dispos.* **31**, 670–676 (2003).
64. Wilson, W. III. *et al.* Characterization of a common deletion polymorphism of the UGT2B17 gene linked to UGT2B15. *Genomics* **84**, 707–714 (2004).
65. Xue, Y. *et al.* Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *Am. J. Hum. Genet.* **78**, 659–670 (2006).
66. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
67. Rademakers, R., Cruts, M. & van Broeckhoven, C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum. Mutat.* **24**, 277–295 (2004).
68. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37**, 129–137 (2005).
69. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
70. Khaja, R. *et al.* Genome assembly comparison identifies structural variants in the human genome. *Nature Genet.* **38**, 1413–1418 (2006).
71. Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S. & Devine, S. E. Natural genetic variation caused by transposable elements in humans. *Genetics* **168**, 933–951 (2004).

72. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: an Introduction to Cluster Analysis* (Wiley, New York, 1990).
73. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
74. Walsh, E. C. *et al.* Searching for signals of evolutionary selection in 168 genes related to immune function. *Hum. Genet.* **119**, 92–102 (2006).
75. Dermitzakis, E. T. *et al.* Comparison of human chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. *Genome Res.* **14**, 852–859 (2004).
76. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2**, 360–369 (2001).
77. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Altshuler, C. Bird, Y. Chen, M. Daly, C. Fahey, A. M. Joseph-George, Y. He, K. Hirose, Z. Hu, V. Jayanth, C. Langford, M. Li, C. Lu, G. Liu, Z. Liu, H. Meguro, L. Pantano, T. Paton, I. Pe'er, S. Pullenayegum, Y. Qi, S. Russell, M. Schachowsky, M. Shago, K. Shiina and Y. Xue for advice, sharing data, cell lines, technical assistance or bioinformatics support. The Centre for Applied Genomics at The Hospital for Sick Children and the Microarray Facility of the Wellcome Trust Sanger Institute are acknowledged for database support and array printing, respectively. We also thank J. R. Lupski and J. Pritchard for comments on earlier versions of the manuscript. The research was supported by The Wellcome Trust (M.E.H., N.P.C., C.T.S.), Canada Foundation of Innovation and Ontario

Innovation Trust (S.W.S.), Canadian Institutes of Health Research (CIHR) (S.W.S.), Genome Canada/Ontario Genomics Institute (S.W.S.), the McLaughlin Centre for Molecular Medicine (S.W.S.), Ontario Ministry of Research and Innovation (S.W.S.), the Hospital for Sick Children Foundation (S.W.S.), The Department of Pathology, Brigham and Women's Hospital (C.L.), The Leukemia and Lymphoma Society (C.L.), the Core Research for Evolutional Science and Technology (CREST) from the Japan Science and Technology Agency, Grant-in-Aid for Scientific Research (H.A.), Grants-in-Aid for Young Scientists and Scientific Research on Priority Areas 'Applied Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan (S.I.), the 'Departament d'Universitats Recerca i Societat de la Informació' (X.E.) and the Genoma España and Genome Canada joint R+D+I projects (X.E. and S.W.S.), the Generalitat de Catalunya (Departament d'Universitats and Departament de Salut), the 'Fondo de Investigaciones Sanitarias de la Seguridad Social, Instituto de Salud Carlos III', and the National Genotyping Center Barcelona Node (CeGen) supported by Genoma España (X.E.), and a Packard Fellowship to J. K. Pritchard (D.C.). L.F. is supported by a fellowship from CIHR; R.R. by a Sanger Institute Postdoctoral Fellowship; and A.C. from the Natural Science and Engineering Research Council. S.W.S. is an Investigator of the CIHR and International Scholar of the Howard Hughes Medical Institute.

Author Contributions R.R., S.I., K.R.F., L.F. and G.H.P. are first authors with equal contribution; C.T.-S., N.P.C., H.A., C.L., K.W.J., S.W.S. and M.E.H. are principal investigators with equal contribution.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to M.E.H. (meh@sanger.ac.uk) or S.W.S. (steve@genet.sickkids.on.ca).

Real-time observation of trigger factor function on translating ribosomes

Christian M. Kaiser¹, Hung-Chun Chang¹, Vishwas R. Agashe^{1†}, Sathish K. Lakshmipathy¹, Stephanie A. Etchells¹, Manajit Hayer-Hartl¹, F. Ulrich Hartl¹ & José M. Barral^{1,2}

The contribution of co-translational chaperone functions to protein folding is poorly understood. Ribosome-associated trigger factor (TF) is the first molecular chaperone encountered by nascent polypeptides in bacteria. Here we show, using fluorescence spectroscopy to monitor TF function and structural rearrangements in real time, that TF interacts with ribosomes and translating polypeptides in a dynamic reaction cycle. Ribosome binding stabilizes TF in an open, activated conformation. Activated TF departs from the ribosome after a mean residence time of ~10 s, but may remain associated with the elongating nascent chain for up to 35 s, allowing entry of a new TF molecule at the ribosome docking site. The duration of nascent-chain interaction correlates with the occurrence of hydrophobic motifs in translating polypeptides, reflecting a high aggregation propensity. These findings can explain how TF prevents misfolding events during translation and may provide a paradigm for the regulation of nucleotide-independent chaperones.

Most proteins must fold into precise three-dimensional structures to become biologically active. In the crowded cytosol, molecular chaperones promote efficient folding by transiently shielding hydrophobic regions of non-native polypeptides, preventing misfolding and aggregation^{1–3}. This process begins during translation, but the mechanisms underlying chaperone–nascent-chain interactions are poorly understood. In bacteria, the chaperone trigger factor (TF, 48 kDa) binds to the ribosome at the polypeptide exit site through interactions with ribosomal proteins L23 and L29 and the 23S RNA^{4–6}, and contacts the emerging polypeptide^{7,8}. The ATP-dependent Hsp70 homologue DnaK interacts with nascent polypeptides subsequently to TF⁹. *Escherichia coli* can tolerate the deletion of either chaperone at 37 °C, but a combined deletion is lethal at temperatures above 30 °C, apparently owing to large-scale protein aggregation^{1–3,9}. TF and DnaK in cooperation not only increase the efficiency of *de novo* protein folding, they can also affect the folding mechanism by delaying folding relative to translation¹⁰. This effect may contribute to the often inefficient folding of eukaryotic multidomain proteins in bacteria^{10,11}. TF is absent in eukaryotes, and thus represents a chaperone paradigm adjusted to the faster translation speed and smaller average protein size in bacteria.

Unlike DnaK, TF does not bind nucleotide, raising the question of how its chaperone activity is regulated. The concentration of TF exceeds that of ribosomes (~50 μM and ~30 μM, respectively)^{12,13}. TF binds free ribosomal 50S subunits with a K_d of ~1 μM (ref. 14) and is probably present in stoichiometric complexes with ribosomes *in vivo*. Purified TF forms dimers with a K_d of 1–2 μM (ref. 14). Thus, non-ribosome-bound TF is expected to exist as a dimer. X-ray crystallography studies have revealed a three-domain structure of TF^{15,16} (Fig. 1a). The amino-terminal domain mediates ribosome interaction and may contribute to substrate binding. A peptidyl-prolyl *cis/trans* isomerase (PPIase) domain resides at the opposite end of the molecule. The significance of this domain has remained enigmatic, as PPIase activity is dispensable for function^{17–19}. The carboxy-terminal domain lies in the middle of the molecule, forming two tips. The

structure of the N-terminal domain of *E. coli* TF in complex with the large ribosomal subunit of the archaeon *Haloarcula marismortui*, which lacks a TF homologue, suggested a model of the position of full-length TF on the ribosome¹⁵. A similar orientation was shown for the N-terminal domain of *Deinococcus radiodurans* TF in complex with its cognate 50S subunit^{5,6}. The model predicts that the C-terminal tips and a segment of the N-terminal domain constitute a crevice that may interact with polypeptides emerging from the exit tunnel. It was also suggested that ribosome-bound TF provides a protected space over the polypeptide exit site for small proteins to fold¹⁵. However, this space is diminished in the complex of TF with the bacterial ribosome owing to the presence of a long loop protruding from eubacterial ribosomal protein L24 (ref. 5).

Here we show that TF interacts with ribosomes and translating polypeptides in a dynamic reaction cycle involving all three functional domains. Binding to the ribosome conformationally activates TF for nascent-chain association. Activated TF departs from the ribosome but may remain bound to the elongating polypeptide. The duration of this interaction is prolonged with nascent chains exposing motifs of high mean hydrophobicity, thus explaining how TF prevents aggregation and delays protein folding relative to translation¹⁰.

Structural rearrangement of TF upon ribosome binding

Fluorescence resonance energy transfer (FRET) allows for the analysis of dynamic conformational changes within protein molecules^{20,21}. We incorporated exogenous fluorophores at specific sites in *E. coli* TF (Fig. 1a) to investigate its interactions with ribosomes and nascent polypeptides. TF was labelled with a coumarin (C) derivative as a donor fluorophore and a fluorescein (F) derivative as an acceptor fluorophore (TF–CF) (see Methods). TF labelled with coumarin alone (TF–C) served as a donor-only control (Supplementary Information). Labelled TF molecules were similar to wild type in their ability to increase the efficiency of protein folding upon *in vitro* translation¹⁰ and their K_d for self-association and ribosome binding

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. ²Department of Neuroscience and Cell Biology, The University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0620, USA. †Present address: Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK.

(Supplementary Fig. 1). Notable intramolecular FRET efficiencies (E_{FRET}) were observed with TF differentially labelled at various positions at concentrations favouring monomers (Fig. 1b).

Upon ribosome binding, the N-terminal domain of *D. radiodurans* TF adopts a conformation different from that in free TF⁶. We investigated conformational rearrangements with labelled full-length TF under conditions that would favour ~70% ribosome binding (Supplementary Information) and found that E_{FRET} was markedly reduced compared with unbound TF (Fig. 1b, middle panel). As the presence of ribosomes did not affect the Förster distance of the donor–acceptor pair used (data not shown), these measurements indicate an increase in distance between the TF-coupled fluorophores upon ribosome binding (Fig. 1b, middle panel; Supplementary Information).

The possible arrangement of TF protomers in the dimer was investigated by intermolecular FRET between TF molecules singly labelled at various positions. E_{FRET} at concentrations favouring dimers depended on fluorophore position (Fig. 1c). Addition of excess unlabelled TF led to the disappearance of FRET with a $t_{1/2}$ value of ~1 s (Supplementary Fig. 2), indicating rapid dimer dissociation kinetics. Computations of the possible configurations consistent with the E_{FRET} values²² converged into a model in which the proposed substrate-binding crevice of TF may be relatively occluded in the dimer (Supplementary Fig. 2). E_{FRET} decreased from 0.4 to 0.16 upon addition of ribosomes at a concentration favouring ~60% TF binding, confirming that the TF dimer is not preserved upon ribosome binding^{15,23} (Fig. 1c). These results suggest that binding to the ribosome stabilizes TF monomers and induces conformational aperture, rendering TF competent to receive the nascent polypeptide chain.

Recruitment of TF to translating ribosomes

Fluorescence emission of TF labelled with BADAN (TF-B; see Methods) at position 14 decreases upon ribosome binding¹⁴. To monitor whether TF is recruited to ribosomes actively engaged in translation, we used an *E. coli* reconstituted cell-free protein synthesis system containing 250 nM ribosomes and negligible amounts of TF²⁴ (Supplementary Fig. 3). TF-B was added to the translation system at a concentration of 250 nM, resulting in ~17% TF-ribosome complexes. Transcription–translation was initiated by adding a plasmid encoding the previously characterized TF substrate firefly luciferase (FL)¹⁰, and TF-B fluorescence emission was monitored in real time. BADAN emission decreased after a lag phase of ~3 min and reached a stable level corresponding to a ~30% decrease after ~25 min (Fig. 2a, top panel). The fluorescence signal returned to baseline

when ribosome–nascent-chain complexes were dissociated with EDTA or RNase A (data not shown). ³⁵S-methionine labelling revealed that full-length FL chains appeared after ~20 min, and their amount increased linearly by the time the BADAN emission reached a stable level (Fig. 2a, top panel). Linear FL accumulation progressed beyond 60 min. The BADAN-labelled FRK/AAA TF mutant, which is unable to bind ribosomes⁴, did not show such fluorescence changes (Fig. 2a, top panel). Thus, the decrease in BADAN fluorescence reflects increased recruitment of TF-B to translating ribosomes, as confirmed by direct binding analysis (Supplementary Fig. 3). Once all of the available ribosomes in the system are translating, TF-B recruitment and protein synthesis reach a steady state, reflected by constant fluorescence and protein accumulation.

Next, we measured changes in TF intramolecular E_{FRET} as a result of translation with TF–CF labelled at positions 14 and 150 or 14 and 326. TF labelled with coumarin alone (TF–C) served as a donor-only control. E_{FRET} decreased with kinetics very similar to those of TF–B fluorescence (Fig. 2a, middle panel). Thus, the equilibrium between compact and open TF molecules is shifted towards the open conformation as ribosomes become engaged in translation. Steady-state fluorescence anisotropy of TF–C confirmed that TF is shifted towards the ribosome-bound form as a result of translation (Fig. 2a, bottom panel). Binding analyses demonstrated that translation led to a ~30-fold reduction in the apparent K_d value of TF for ribosome association (Supplementary Information). This increase in TF affinity for the ribosome is dependent on the emergence of the nascent chain from the exit tunnel (see Fig. 4 below).

Substrate binding can persist after TF–ribosome dissociation

In competition experiments, dissociation of TF–B from complexes with non-translating ribosomal subunits occurred with a $t_{1/2}$ of 10.3 ± 2.2 s ($\pm$ s.d.) at 30 °C (Fig. 2b, top panel), in good agreement with published results¹⁴. Upon the addition of excess unlabelled TF to reactions containing equilibrium complexes of vacant ribosomal subunits and the TF–CF molecules used for FRET measurements, E_{FRET} returned to baseline with a $t_{1/2}$ of 8.6 ± 2.3 s, similar to the kinetics observed for ribosome dissociation (Fig. 2b, top panel). Additionally, the decrease in anisotropy of TF–C upon competition with unlabelled TF occurred with a $t_{1/2}$ value of ~10 s (data not shown). Thus, TF dissociation from vacant ribosomal subunits and intramolecular compaction occur concurrently.

To determine whether ongoing translation would alter the rates of TF conformational compaction and ribosome departure, competition experiments were performed with the reconstituted system

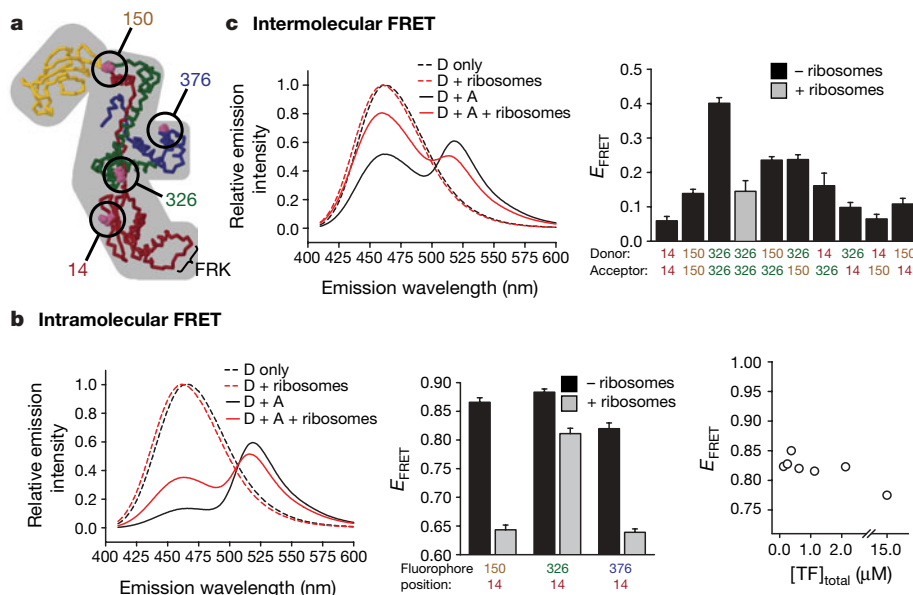


Figure 1 | TF undergoes structural rearrangements upon binding to ribosomes.

a, Crystal structure of TF indicating the positions of exogenous chromophores (pink), FRK residues, and subdomain arrangement¹⁵: ribosome-binding domain (red), PPIase domain (yellow), tip 1 (green) and tip 2 (blue). **b**, Left panel, intramolecular FRET (14–150 pair). Middle panel, intramolecular E_{FRET} ($\pm$ ribosomes) ($n = 5$). Right panel, intramolecular E_{FRET} remains unchanged upon addition of unlabelled TF. Labelled TF was at 250 nM. $[\text{TF}]_{\text{total}}$ total concentration of TF. **c**, Left panel, intermolecular FRET between positions 326 ($\pm$ ribosomes). D, donor; A, acceptor. Right panel, intermolecular E_{FRET} ($\pm$ ribosomes) ($n = 4$). Labelled TF was at 250 nM (donor) and 1 μ M (acceptor). Error bars are $\pm$ s.d.

translating FL. Excess unlabelled TF was added when translation had reached steady state. Ongoing translation did not affect the kinetics of TF dissociation from the ribosome, as measured by TF–B fluorescence ($t_{1/2}$ 11.2 ± 0.9 s) (Fig. 2b, middle panel). In contrast, conformational compaction (TF–CF) showed considerably slower kinetics, with a $t_{1/2}$ of 36.6 ± 3.5 s (Fig. 2b, middle panel). Similarly, dissociation of TF from translating ribosomes measured by anisotropy with TF–C occurred with a $t_{1/2}$ of ~ 45 s (data not shown). Thus, upon departure from the ribosome docking site, TF remains tethered to the ribosome-bound nascent chain. A new TF molecule may dock onto the now vacant ribosome, consistent with biochemical evidence that more than one TF molecule can associate per ribosome–nascent-chain complex¹⁰. Notably, FRK/AAA TF mutant was unable to act as a competitor (Fig. 2b, top and middle panels).

Next, we investigated whether the TF PPIase domain, which can bind unfolded polypeptides *in vitro*²⁵, is involved in nascent-chain association by using a TF version entirely lacking the PPIase domain (Δ PPIase; see Methods). BADAN-labelled Δ PPIase (Δ PPIase–B) exhibited ribosome binding kinetics very similar to those of TF during FL translation, and Δ PPIase was as effective as TF in competition experiments (data not shown). However, when labelled Δ PPIase was displaced by excess unlabelled TF in a FL translation reaction, intramolecular compaction and ribosome dissociation occurred with $t_{1/2}$ values of ~ 23 s and ~ 8 s, respectively (data not shown). Thus, Δ PPIase undergoes significantly faster compaction than TF ($t_{1/2}$ of ~ 37 s). These findings indicated that the PPIase domain may mediate an additional binding event following initial TF interaction with the ribosome–nascent-chain complex. Indeed, crosslinking experiments with TF molecules having the photocrosslinker *para*-benzoyl-L-phenylalanine²⁶ incorporated at position 320 in tip 1 of the C-terminal domain or at residue 233 in the PPIase domain demonstrated that both domains bind to the nascent chain (Supplementary

Fig. 4). Crosslinking with the PPIase domain was observed with FL nascent chains comprising residues 1–164 but not with short chains of residues 1–77, although both chain lengths crosslinked efficiently with tip 1.

The above results indicated that TF may remain bound to substrate proteins post-translationally. A post-translational state was mimicked by disrupting complexes of ribosome-stalled FL nascent chains by EDTA treatment²⁷. Upon the addition of EDTA, TF dissociation from the ribosome occurred with a $t_{1/2}$ of 14.1 ± 0.1 s, whereas TF molecular compaction occurred more slowly, with a $t_{1/2}$ of 21.2 ± 1.1 s (Fig. 2b, bottom panel). Thus, TF can remain associated with nascent chains even after their departure from the ribosome but is unable to rebind them, as this function is mediated by ribosome docking.

Hydrophobic motifs in nascent chains determine TF binding

DnaK binds preferentially to motifs composed of four or five hydrophobic residues flanked by basic residues²⁸, whereas the TF PPIase domain interacts with stretches of eight residues enriched in basic and aromatic side chains²⁵. To address whether TF discriminates among different nascent polypeptides, a range of eukaryotic and prokaryotic proteins were analysed in translation reactions in competition experiments as described above (Fig. 3a, b). Ribosome departure kinetics were similar among the polypeptides examined (Fig. 3a, inset). However, as was found for FL, molecular compaction occurred more slowly than ribosome release for *E. coli* GatD as well as for OxyR and HcaD (see below) (Fig. 3a). In contrast, TF dissociation from ribosomes translating *E. coli* GatZ, RecA and MetK occurred at a rate similar to that of compaction (Fig. 3a). Therefore, TF remains bound to certain polypeptides for varying lengths of time after leaving the ribosome, whereas nascent-chain release and ribosome departure occur concomitantly for others.

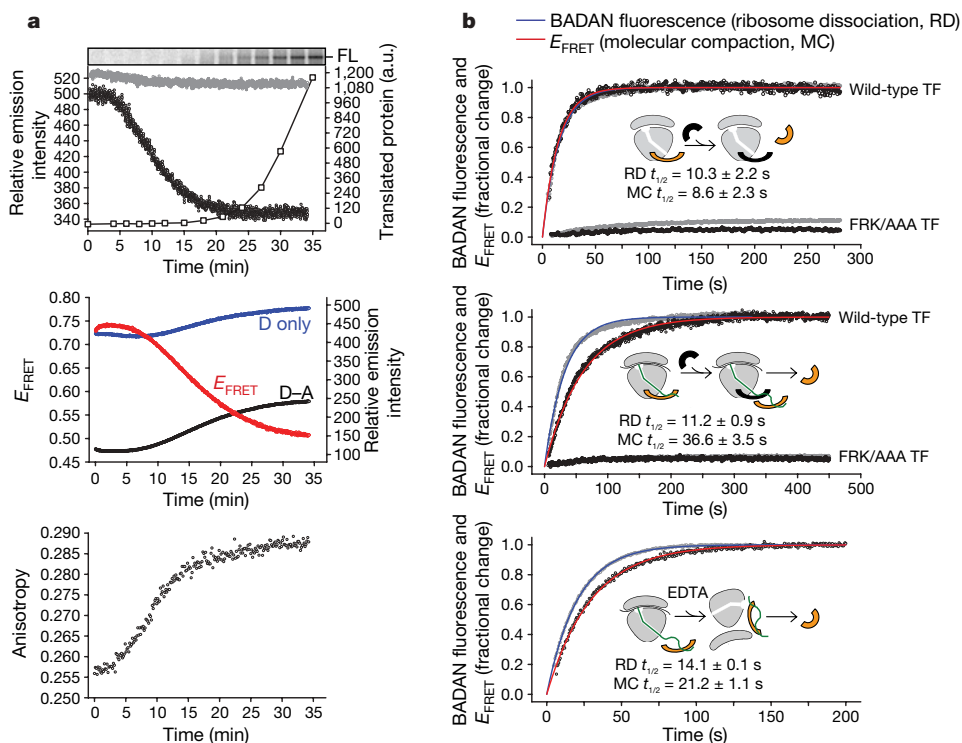


Figure 2 | TF remains associated with the nascent chain upon leaving the ribosome. **a**, Real-time measurements during FL translation. Top panel, fluorescence emission of wild-type TF (black circles) and the FRK/AAA mutant (grey circles) labelled with BADAN, and FL accumulation (black squares). Above gel shows ³⁵S-Met-labelled FL protein (fluorography). a.u., arbitrary units. Middle panel, donor fluorescence of TF–C (blue) and TF–CF (positions 14 and 150; black), and E_{FRET} (red). Bottom panel, fluorescence

anisotropy of TF–C. **b**, Fractional change in fluorescence emission (blue) and E_{FRET} (red) of TF–B and TF–CF (orange molecules) in reactions of: top panel, vacant ribosomes upon competition with unlabelled TF (black molecules) or TF FRK/AAA; middle panel, ribosomes translating FL upon competition with unlabelled TF as in panel above; bottom panel, stalled full-length FL nascent chains (green molecules) released with EDTA.

The kinetics of TF molecular compaction correlated with the occurrence of motifs of high mean hydrophobicity²⁹ in nascent chains (Fig. 3b) (Supplementary Information). An optimal correlation was obtained with TF interaction motifs of at least 15 consecutive residues (Supplementary Information), which can be accommodated within the ~65-Å-long putative substrate-binding crevice of TF^{15,30}. Analysis of the cytosolic proteome of *E. coli* predicted that 243 proteins (~15%) may exhibit prolonged interaction with TF (Supplementary Fig. 6). Indeed, association of TF with OxyR and HcaD, which were selected as candidate proteins containing multiple putative TF interaction motifs, was found to be prolonged, highlighting the predictive power of the established correlation.

FL is composed of two structural domains³¹. Predicted TF interaction motifs cluster in two regions within the N-terminal domain, whereas the C-terminal domain contained no such motifs (Fig. 3c). We translated an N-terminal segment (N1; residues 1–224), a middle fragment (N2; residues 225–436) and the complete C-terminal domain (C; residues 437–544) of FL (Fig. 3c). Only fragments N1 and N2, containing proposed TF interaction motifs, led to slower rates of TF molecular compaction (Fig. 3d, left and middle panels). Small domains lacking contiguous sequences with high mean hydrophobicity, such as the FL C-terminal domain, are expected to fold rapidly with little danger of aggregating. Thus, the duration of TF–substrate interaction would correlate with the aggregation propensity of the substrate.

Burial of hydrophobic sequences results in TF displacement

The titin I27 domain acquires a compact β -sandwich structure^{32,33} co-translationally if allowed to fully emerge from the ribosome (data not shown). We analysed TF–B recruitment to I27 nascent chains stalled at different positions throughout their lengths (Supplementary Information and Supplementary Fig. 5). I27 nascent chains stalled at position 25 reside entirely within the ~100-Å-long ribosomal exit tunnel^{34,35}, because 25 residues in an extended conformation³⁰ would span only ~85 Å. No additional TF recruitment beyond that measured for free ribosomal subunits was observed with the 25-residue chains, indicating that the process of translation *per se* is insufficient to procure high-affinity binding of TF to the ribosome. Similarly, I27 nascent chains of 38 or 58 amino acids showed no TF recruitment. Owing to the presence of a His₆ sequence, the extreme N terminus of the I27 version used is polar and therefore should not

interact strongly with TF. However, a more hydrophobic region is present between residues 17 and 40 (Fig. 4a), which would be expected to lie well outside the ribosome when I27 is ~80 residues long. Notable TF recruitment was observed with I27 stalled at position 78, and a further increase was observed when I27 was stalled at its last codon (full length), which coincides with the emergence of a second hydrophobic stretch (Fig. 4b). This result is consistent with TF containing two binding sites for nascent chains.

Because the folding of I27 is highly cooperative³³, segments containing non-full-length I27 polypeptides outside the ribosomal exit tunnel would not be expected to fold. However, when full-length I27 was allowed to emerge from the ribosome by stalling nascent chains with extensions composed of 30 to 50 residues of the repeat (GGGS)_n, TF recruitment decreased gradually (Fig. 4b). TF preserved an affinity for I27 even after the full-length protein had emerged from the ribosome (full-length plus 40–50 extra residues) (Fig. 4b), indicating that delayed folding is a fundamental consequence of TF action, even with domains that fold in a highly cooperative manner. Sedimentation analysis confirmed the recruitment of TF to ribosome–nascent-chain complexes (Supplementary Fig. 3). These results demonstrate that the interaction of TF with translating ribosomes is stabilized by nascent chains containing exposed hydrophobic segments. However, concomitantly with the burial of these hydrophobic regions due to folding, TF is released from its substrate. Polypeptides that bury hydrophobic residues slowly may interact with TF for prolonged periods.

Implications for mechanism

We have used fluorescence spectroscopy to examine the action of TF on a variety of polypeptide chains during their biosynthesis. These experiments have revealed an exquisitely regulated interaction cycle of TF with ribosome–nascent-chain complexes (Fig. 4c). TF is in rapid dimer–monomer equilibrium in solution (Fig. 4c, reaction 1). Ribosomes withdraw TF from the rapidly replenished pool of monomers (Fig. 4c, reaction 2). Upon binding at the ribosomal exit site, the TF monomer undergoes a conformational expansion, allowing subdomain alignment for association with the emerging polypeptide. In the absence of translation, TF performs a default binding cycle with a mean residence time on the ribosome of ~10 s and relaxes to its compact conformation concomitantly. During translation, ribosome-bound TF associates with its nascent-chain

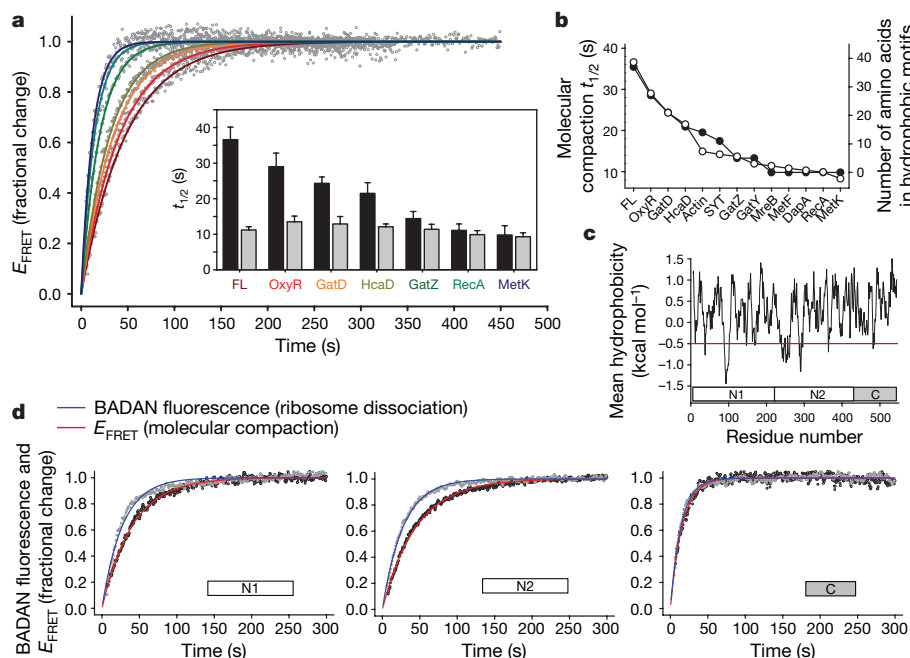


Figure 3 | The nature of the nascent polypeptide influences its interaction with TF. **a**, Kinetics of TF compaction upon departure from the ribosome during translation. Fractional change in maximum E_{FRET} of reactions containing TF–CF (14–326) upon competition with unlabelled TF. Inset, $t_{1/2}$ values for molecular compaction (black bars) and ribosome dissociation (grey bars) ($n = 5$). **b**, Correlation between the extent of motifs of high hydrophobicity, given as number of amino acids (filled circles) and $t_{1/2}$ values of TF molecular compaction (open circles). **c**, Regions of high hydrophobicity (below red line) along the sequence of the N- (N1 and N2) and C- (C) terminal FL domains. **d**, Kinetics of TF–ribosome dissociation (blue) and TF compaction (red) of reactions translating the N1 (left panel), N2 (middle panel) and C-terminal (right panel) domain fragments of FL given as fractional changes in BADAN fluorescence emission and E_{FRET} . Error bars are $\pm$ s.d.

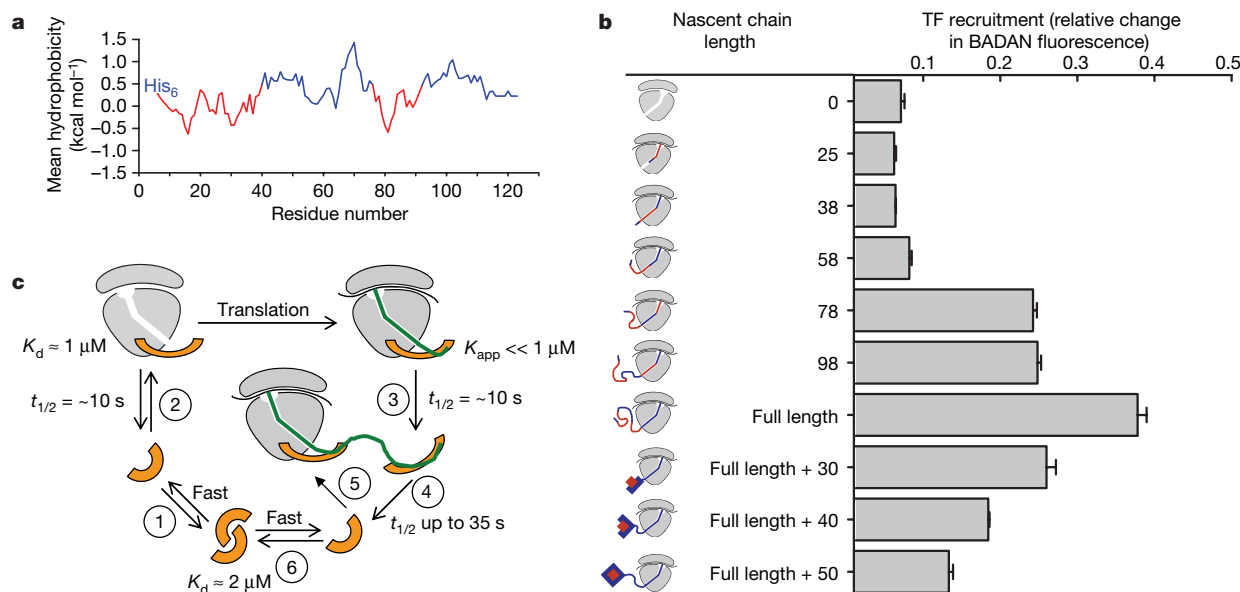


Figure 4 | Chain-length dependence of TF interaction with titin I27 and model of TF function. **a**, Regions of high (red) and low (blue) hydrophobicity along the I27 sequence. **b**, Recruitment of TF-B upon translation of ribosome-tethered I27 polypeptides stalled at the indicated positions ($n = 5$), with diagrams on the left coloured as in **a**. Three-dimensional structure acquisition is represented schematically by the diamond shapes. Error bars are $\pm$ s.d. **c**, Model of TF function. Free TF is in rapid monomer–dimer equilibrium (reaction 1). TF binding to the ribosome results in TF activation by conformational expansion and monomer

stabilization (reaction 2). During translation, TF binds to hydrophobic segments of the emerging nascent chain. Substrate-bound TF leaves the ribosome but preserves its expanded state (reaction 3) remaining substrate-bound for varying periods of time (reaction 4). With ongoing translation, a second TF molecule may dock onto the now vacated ribosome and bind to the elongating polypeptide (reaction 5). TF dissociates from the nascent chain, resumes its compact conformation and becomes available to replenish the dimer pool (reaction 6).

substrate. Although dissociation from the ribosome docking site remains essentially unchanged (Fig. 4c, reaction 3), substrate-bound TF preserves its open conformation and its association with the nascent chain for varying periods of time (up to $>30 \text{ s}$) upon leaving the ribosome (Fig. 4c, reaction 4). With longer chains, an additional TF molecule can then dock onto the now vacated ribosome and associate with the elongating polypeptide (Fig. 4c, reaction 5). Bound TF molecules eventually dissociate from their polypeptide substrate, resuming their compact conformation. Free TF dimerizes, thereby (partially) occluding its substrate-binding regions (Fig. 4c, reaction 6).

The salient feature of this mechanism is the cycling of TF between a dimer form in free solution and a conformationally activated monomer that is induced by ribosome docking, coupling the step of substrate association to ribosome binding. Default cycling provides sufficient time for TF to scan nascent chains for TF interaction motifs. As TF and the signal recognition particle (SRP) may bind simultaneously to the ribosome, SRP would be able to preferentially recognize the signal anchor sequences of nascent inner-membrane proteins^{36–41}. During the average time window of TF binding to the ribosome ($\sim 10 \text{ s}$), approximately 100–200 amino acids are being synthesized¹², allowing completion of synthesis for small proteins. As the residence time of TF on the ribosome is not dramatically altered by translation, TF can depart from the ribosome but maintain its association with longer nascent chains, vacating the ribosome docking site for a new TF monomer.

How does this mechanism serve to support protein folding? Our data indicate that the main role of TF is to shield hydrophobic sequence elements in unfolded nascent chains, thereby preventing misfolding and aggregation. A primary binding site composed of the N- and C-terminal domains may function together with a secondary site in the PPIase domain. Small, cooperatively folding proteins would reach native state rapidly upon TF dissociation. Interference by TF with post-translational completion of folding is avoided because efficient TF binding to protein substrate occurs only in the vicinity of the ribosome. The secondary binding site may be used

preferentially by polypeptides with complex structures that expose multiple hydrophobic stretches during translation and fail to bury these segments rapidly upon TF cycling. Such proteins may be particularly aggregation-sensitive⁴² and require prolonged protection by TF or transfer to downstream chaperones for folding.

METHODS

Preparation of TF variants. Single-cysteine mutants R14C, T150C, E326C and S376C, and double-cysteine mutants R14C-T150C, R14C-E326C and R14C-S376C, were generated for wild-type *E. coli* TF, as well as the mutant F44A-R45A-K46A⁴, and a mutant in which the PPIase domain (residues 151–243) was substituted with GTSAAAG (Δ PPIase). Dyes for site-specific labelling were 6-bromoacetyl-2-dimethyl-aminonaphthalene (BADAN), 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM), and fluorescein-5-maleimide. See Supplementary Information for protein expression, purification and labelling methods.

Equilibrium fluorescence measurements. Methods and related calculations are described in Supplementary Information. TF-B or Δ PPIase-B measurements were performed with 250 nM labelled protein, excited at 397 nm and emission spectra were collected from 410 to 600 nm. Intermolecular FRET measurements were performed with 250 nM donor-labelled TF and 1 μM acceptor-labelled TF, or 1 μM unlabelled TF instead of acceptor-labelled for donor-only measurements. The donor was excited at 397 nm, and emission was recorded from 410 to 650 nm. Purified ribosomes⁴³ were added as indicated. Intramolecular FRET measurements were performed in a similar fashion, except that the samples contained 250 nM of either donor-only or donor- and acceptor-labelled TF (or the respective Δ PPIase versions), and unlabelled TF and ribosomes at the specified concentrations, where indicated.

Real-time fluorescence measurements during translation. Transcription–translation was performed in *E. coli* PURESYSYSTEM classic II (Post Genome Institute Co.) according to the manufacturer's instructions with template plasmids^{10,44} or PCR products lacking stop codons. See Supplementary Information for details.

Competition experiments. For displacement experiments from vacant ribosomal subunits, 250 nM labelled TF were mixed with 2 μM ribosomes. For displacement experiments with translating ribosomes, translation in the presence of 250 nM labelled TF was allowed to progress until fluorescence and protein production reached a steady state (35 min). Unlabelled competitor

was added to a final concentration of 20 μ M and fluorescence emission was measured over time.

Received 14 March; accepted 30 August 2006.

Published online 15 October 2006.

1. Frydman, J. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. *Annu. Rev. Biochem.* **70**, 603–647 (2001).
2. Young, J. C., Agashe, V. R., Siegers, K. & Hartl, F. U. Pathways of chaperone-mediated protein folding in the cytosol. *Nature Rev. Mol. Cell Biol.* **5**, 781–791 (2004).
3. Deuerling, E. & Bukau, B. Chaperone-assisted folding of newly synthesized proteins in the cytosol. *Crit. Rev. Biochem. Mol. Biol.* **39**, 261–277 (2004).
4. Kramer, G. *et al.* L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419**, 171–174 (2002).
5. Schlunzen, F. *et al.* The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure (Camb.)* **13**, 1685–1694 (2005).
6. Baram, D. *et al.* Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc. Natl Acad. Sci. USA* **102**, 12017–12022 (2005).
7. Hestekamp, T., Hauser, S., Lutcke, H. & Bukau, B. *Escherichia coli* trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. *Proc. Natl Acad. Sci. USA* **93**, 4437–4441 (1996).
8. Valent, Q. A. *et al.* Early events in preprotein recognition in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. *EMBO J.* **14**, 5494–5505 (1995).
9. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
10. Agashe, V. R. *et al.* Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* **117**, 199–209 (2004).
11. Chang, H. C., Kaiser, C. M., Hartl, F. U. & Barral, J. M. *De novo* folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. *J. Mol. Biol.* **353**, 397–409 (2005).
12. Bremer, H. & Dennis, P. P. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhart, F. C.) 1553–1569 (ASM Press, Washington DC, 1996).
13. Lill, R., Crooke, E., Guthrie, B. & Wickner, W. The “trigger factor cycle” includes ribosomes, presecretory proteins, and the plasma membrane. *Cell* **54**, 1013–1018 (1988).
14. Maier, R., Eckert, B., Scholz, C., Lilie, H. & Schmid, F. X. Interaction of trigger factor with the ribosome. *J. Mol. Biol.* **326**, 585–592 (2003).
15. Ferbitz, L. *et al.* Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).
16. Ludlam, A. V., Moore, B. A. & Xu, Z. The crystal structure of ribosomal chaperone trigger factor from *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **101**, 13436–13441 (2004).
17. Genevax, P. *et al.* *In vivo* analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* **5**, 195–200 (2004).
18. Kramer, G. *et al.* Functional dissection of *Escherichia coli* trigger factor: unraveling the function of individual domains. *J. Bacteriol.* **186**, 3777–3784 (2004).
19. Kramer, G. *et al.* Trigger factor peptidyl-prolyl *cis/trans* isomerase activity is not essential for the folding of cytosolic proteins in *Escherichia coli*. *J. Biol. Chem.* **279**, 14165–14170 (2004).
20. Stryer, L. Fluorescence spectroscopy of proteins. *Science* **162**, 526–533 (1968).
21. Woolhead, C. A., McCormick, P. J. & Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725–736 (2004).
22. Schroder, G. F. & Grubmüller, H. FRETsg: biomolecular structure model building from multiple FRET experiments. *Comp. Phys. Comm.* **158**, 150–157 (2004).
23. Patzelt, H. *et al.* Three-state equilibrium of *Escherichia coli* trigger factor. *Biol. Chem.* **383**, 1611–1619 (2002).
24. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature Biotechnol.* **19**, 751–755 (2001).
25. Patzelt, H. *et al.* Binding specificity of *Escherichia coli* trigger factor. *Proc. Natl Acad. Sci. USA* **98**, 14244–14249 (2001).
26. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 11020–11024 (2002).
27. Michalski, C. J., Sells, B. H. & Morrison, M. Molecular morphology of ribosomes. Localization of ribosomal proteins in 50-S subunits. *Eur. J. Biochem.* **33**, 481–485 (1973).
28. Rudiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **16**, 1501–1507 (1997).
29. Roseman, M. A. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* **200**, 513–522 (1988).
30. Creighton, T. E. *Proteins: Structures and Molecular Properties* (W. H. Freeman and Co., New York, 1984).
31. Conti, E., Franks, N. P. & Brick, P. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4**, 287–298 (1996).
32. Improtà, S., Politou, A. S. & Pastore, A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* **4**, 323–337 (1996).
33. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
34. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
35. Harms, J. *et al.* High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679–688 (2001).
36. Ullers, R. S. *et al.* Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the *Escherichia coli* ribosome. *J. Cell Biol.* **161**, 679–684 (2003).
37. Beck, K., Wu, L. F., Brunner, J. & Müller, M. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *EMBO J.* **19**, 134–143 (2000).
38. Eisner, G., Moser, M., Schäfer, U., Beck, K. & Müller, M. Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. *J. Biol. Chem.* **281**, 7172–7179 (2006).
39. Buskiewicz, I. *et al.* Trigger factor binds to ribosome-signal-recognition particle (SRP) complexes and is excluded by binding of the SRP receptor. *Proc. Natl Acad. Sci. USA* **101**, 7902–7906 (2004).
40. Raine, A., Ivanova, N., Wikberg, J. E. & Ehrenberg, M. Simultaneous binding of trigger factor and signal recognition particle to the *E. coli* ribosome. *Biochimie* **86**, 495–500 (2004).
41. Ullers, R. S. *et al.* Sequence-specific interactions of nascent *Escherichia coli* polypeptides with trigger factor and signal recognition particle. *J. Biol. Chem.* **281**, 13999–14005 (2006).
42. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805–808 (2003).
43. Spedding, G. in *Ribosomes and Protein Synthesis: A Practical Approach* (ed. Spedding, G.) 1–29 (Oxford Univ. Press, Oxford, 1990).
44. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Johnson for advice regarding assessment of fluorophore mobility, A. Bracher for help in designing the Δ Plase construct and advice during building of the TF dimer model, M. Kerner for his help with bioinformatics analyses, R. Boteva and K. Chakroborty for discussion, P. Genevax and C. Georgopoulos for the GP367 *E. coli* strain and P. Schultz for reagents for *in vivo* crosslinking. J.M.B. was supported by a fellowship from the International Human Frontier Science Program Organization. Support by the Ernst Jung Foundation and the European Union is also acknowledged.

Author Contributions F.U.H. was the project leader. C.M.K. and J.M.B. were responsible for experimental and project design. C.M.K. performed most of the experiments. V.R.A. made conceptual contributions and H.-C.C., V.R.A., S.K.L., S.A.E., M.H.-H. and J.M.B. performed some of the experiments. F.U.H. and J.M.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.U.H. (uhartl@biochem.mpg.de) or J.M.B. (jmbarral@utmb.edu).

Magnetic vortex core reversal by excitation with short bursts of an alternating field

B. Van Waeyenberge^{1,2}, A. Puzic¹, H. Stoll¹, K. W. Chou¹, T. Tyliczszak³, R. Hertel⁴, M. Fähnle¹, H. Brückl^{5†}, K. Rott⁵, G. Reiss⁵, I. Neudecker⁶, D. Weiss⁶, C. H. Back⁶ & G. Schütz¹

The vortex state, characterized by a curling magnetization, is one of the equilibrium configurations of soft magnetic materials^{1–4} and occurs in thin ferromagnetic square and disk-shaped elements of micrometre size and below. The interplay between the magneto-static and the exchange energy favours an in-plane, closed flux domain structure. This curling magnetization turns out of the plane at the centre of the vortex structure, in an area with a radius of about 10 nanometres—the vortex core^{5–7}. The vortex state has a specific excitation mode: the in-plane gyration of the vortex structure about its equilibrium position^{8–10}. The sense of gyration is determined by the vortex core polarization¹¹. Here we report on the controlled manipulation of the vortex core polarization by excitation with small bursts of an alternating magnetic field. The vortex motion was imaged by time-resolved scanning transmission X-ray microscopy¹². We demonstrate that the sense of gyration of the vortex structure can be reversed by applying short bursts of the sinusoidal excitation field with amplitude of about 1.5 mT. This reversal unambiguously indicates a switching of the out-of-plane core polarization. The observed switching mechanism, which can be understood in the framework of micromagnetic theory, gives insights into basic magnetization dynamics and their possible application in data storage.

The continuous downscaling in microfabrication technology has enabled the creation of magnetic microstructures and nanostructures with defined sizes and shapes. These structures are currently not only implemented in applications such as data storage and non-volatile magnetic random access memory (MRAM), but also form an interesting playground for the fundamental studies of magnetism on a microscopic level.

In thin film structures, in which the magnetostatic interactions usually force the magnetization to lie parallel to the film plane, typical magnetic configurations occur with domain structures that close the magnetic flux. Square patterns have in this case a typical Landau structure with four triangular domains separated by 90° domain walls. The magnetic vortex is located at the centre of this domain structure, where the four domains meet one another. The curling magnetization cannot stay in the plane at the very centre of the vortex structure because the short-range exchange interaction favours a parallel alignment of neighbouring magnetic moments. The magnetization turns perpendicular to the plane in an area with a radius of about 10 nm, in this way forming the vortex core⁷. The direction of the out-of-plane component of the magnetization is defined as the polarization of the vortex core (up or down) and gives, together with the sense of the in-plane flux closure (clockwise or anticlockwise), the ground-state configuration as illustrated in Fig. 1a–c. A magnetic vortex can store two bits of information¹³: the sense of the in-plane

flux closure can be used as an information carrier (Fig. 1a, b)^{14,15}, and the out-of-plane polarization of the magnetic vortex core can also be regarded as ‘0’ or ‘1’ of a bit element (Fig. 1a, c). However, to switch the vortex core polarization, magnetic fields of the order of 0.5 T (refs 16, 17) are needed.

Here we report on experimental studies towards an easy and reproducible switching of the vortex core polarization by low-field excitations. The dynamics in micrometre-sized and square ferromagnetic Permalloy elements with a Landau magnetic ground state were investigated. The structures were excited with an in-plane sinusoidal magnetic field resulting in a gyrotropic movement of the vortex core around the equilibrium position. As already verified in magneto-optical measurements, this in-plane gyrotropic mode is the lowest excitation mode in elements exhibiting a vortex structure (in the frequency range 100 MHz to 1 GHz (refs 18, 19)). A general theory on the dynamics of magnetic domain structures has been introduced previously⁸. The sense of gyration of a vortex structure is given by the gyrocoupling vector $\mathbf{G} = -2\pi q p \hat{z}$, where q is the topological vorticity, p the vortex core polarization, and $\hat{z}$ a unit vector in the out-of-plane direction⁹. The topological vorticity $q = -1$ gives an antivortex structure (Fig. 1d), but only the vortex structure with topological vorticity $q = +1$ appears in the ground-state configuration (Fig. 1a–c). The sense of gyration is therefore determined only by the out-of-plane vortex core polarization p and in particular is independent of the sense of the in-plane flux closure^{8,9,11}. A change in the sense of gyration indicates unambiguously a change in the direction of the vortex core polarization.

The temporal evolution of the gyrotropic motion for a $1.5 \mu\text{m} \times 1.5 \mu\text{m} \times 50 \text{ nm}$ element was imaged by scanning transmission X-ray microscopy. The excitation frequency was set at 250 MHz and is close to the resonance frequency of about 244 MHz derived from micromagnetic simulations. The excitation field is synchronized with the probing flashes of circularly polarized X-rays, allowing us to take snapshots of the magnetic response in a stroboscopic experiment. This technique, combined with the element-specific X-ray magnetic circular dichroism (XMCD) effect²⁰, permits magnetic imaging with high lateral resolution of about 30 nm, and a time resolution of less than 100 ps can be reached. A more detailed view of the sample and stripline setup is given in Fig. 2 and is described in Methods.

A small 0.1 mT continuous sinusoidal excitation field was applied to induce a permanent gyrotropic motion of the Landau structure. This permits the imaging of the vortex gyration and the determination of the vortex core polarization by its gyration sense. A short magnetic field burst of 4 ns (corresponding to a single period of the radiofrequency (RF) field) was applied, superimposed on the

¹Max-Planck-Institut für Metallforschung, 70569 Stuttgart, Germany. ²Department of Subatomic and Radiation Physics, Ghent University, 9000 Gent, Belgium. ³Advanced Light Source, LBNL, 94720 Berkeley, California, USA. ⁴Institut für Festkörperforschung IFF-9 ‘Elektronische Eigenschaften’, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

⁵Fakultät für Physik, Universität Bielefeld, 33615 Bielefeld, Germany. ⁶Institut für Experimentelle und Angewandte Physik, Universität Regensburg, 93040 Regensburg, Germany.

[†]Present address: Division ‘Nano System Technology’, ARCS research GmbH, Tech Gate Vienna, Donau-City-Strasse 1, 1220 Vienna, Austria.

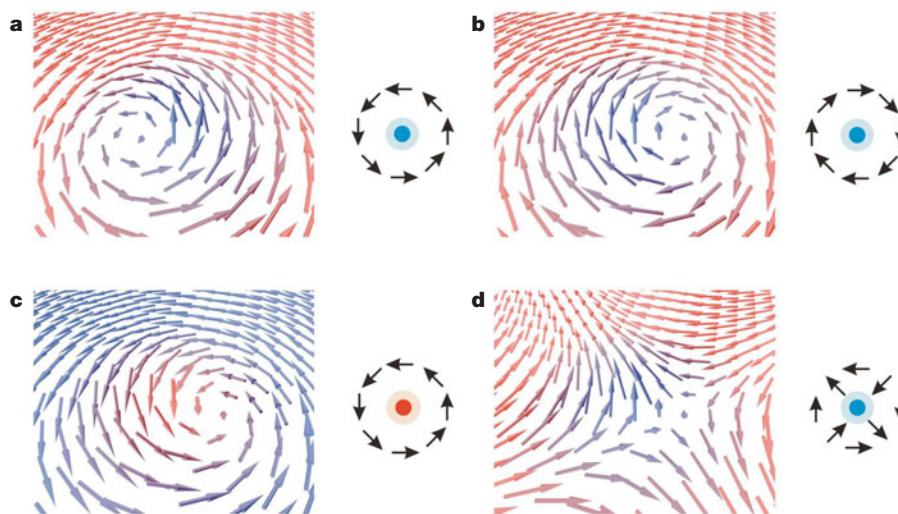


Figure 1 | Three-dimensional and two-dimensional representation of vortex and antivortex structures. Vortex (a, b and c) and antivortex (d) structures are illustrated. In both cases the magnetization turns out of the plane at the centre of the structure—either up (a, b, d) or down (c)—corresponding to the vortex core polarization p . In addition, vortex structures are characterized by an in-plane flux closure, which can be

clockwise (b) or anticlockwise (a, c). A three-dimensional representation is on the left of each panel; a two-dimensional scheme is on the right. The arrows in the two-dimensional schemes represent the in-plane magnetization components; while the coloured dots represent the out-of-plane component (blue, up; red, down).

weak alternating field. The resulting dynamic response of the vortex structure is given in Fig. 3. The position of the vortex core can be followed in the two sequences of images, before and after the burst, taken for equal-phase angle steps with respect to the alternating excitation field. The sense of flux closure of the Landau pattern can be observed directly in the magnetic contrast of the images and remains unchanged during the experiments. The vortex positions at distinct phase angles are extracted from the individual images and depicted on the top part of Fig. 3. Individual positions of the vortex core describe an elliptical trajectory. A change in the sense of gyrotropic vortex motion can be clearly seen before and after applying the burst (Fig. 3, and Supplementary Videos 1 (before the burst) and 2 (after the burst)).

As discussed above, a change in the sense of gyration of the vortex structure is an unambiguous indication of a switching of the vortex core polarization. The field strength of the burst that is necessary to change the direction of the vortex core polarization is found to be about 1.5 mT. It is remarkable that even a short burst of only one monocycle of the alternating field (4 ns) in combination with the

relatively small amplitude of about 1.5 mT is sufficient to toggle the vortex core back and forth.

The vortex core switching observed experimentally has been reproduced by micromagnetic simulations based on the Landau–Lifshitz–Gilbert equation (Supplementary Video 3). A $1.5\ \mu\text{m} \times 1.5\ \mu\text{m} \times 50\ \text{nm}$ structure was excited by an alternating magnetic field at 250 MHz with an amplitude of 0.1 mT. This field induces a small gyration of the vortex core. An additional single burst of the magnetic field with a length of one monocycle (4 ns) and amplitude of 3.5 mT was sufficient to induce switching of the vortex core magnetization. A second identical burst was applied to toggle the vortex core polarization back. These simulations also reveal details of this switching process (Fig. 4, and Supplementary Video 4). A vortex structure with down core polarization (Fig. 4a) is driven by the excitation and a distortion of the out-of-plane vortex structure occurs²¹. Along the vortex core, an area is formed with opposite out-of-plane magnetization with respect to the vortex core (Fig. 4b). The out-of-plane component increases and reaches full out-of-plane orientation (Fig. 4c). When the magnetization continues to turn, the in-plane magnetization changes direction. This is the point at which a vortex–antivortex pair with equal polarizations is created (Fig. 4d, e). The newly formed vortex and antivortex move apart, and the antivortex goes towards the original vortex (Fig. 4f). When the antivortex meets the original vortex, they annihilate with the emission of spin waves²² and only one vortex remains in the structure with an opposite polarization (Fig. 4g, h). The annihilation of a vortex with an antivortex of opposite polarization requires, for topological reasons, the formation of a Bloch point²³. These simulations illustrate some important steps of the reversal of the vortex core polarization, but explaining all the microscopic details will require further investigation.

These experimental results, which can be reproduced by micromagnetic simulations, show that properly tuned bursts of only 4 ns can be used to switch the vortex core. Because the resonance frequency of the gyrotropic mode scales inversely with the lateral dimensions²⁴, much shorter pulses should be sufficient for switching the core polarization in smaller elements. Although their practical realization is still far off, data storage systems based on this core switching scheme could have several advantages, including high thermal stability, insensitivity to external static fields and minimal

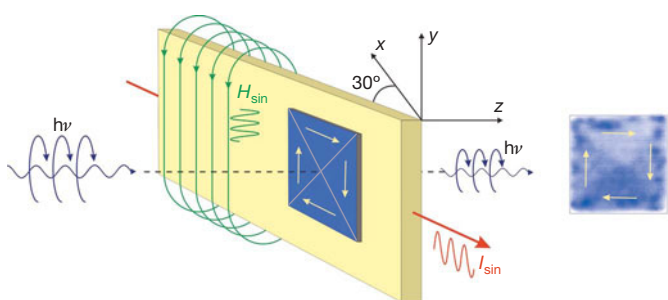


Figure 2 | Schematic overview of the sample and stripline setup. A tilted stripline with a sample is illustrated; an alternating current I_{sin} is sent through, generating an in-plane magnetic field H_{sin} . The circularly polarized X-ray photons are selectively absorbed by the magnetic domains of the sample. An example of the magnetic contrast, proportional to the magnetization along the x direction M_x , is shown at the right. This image of a $1.5\ \mu\text{m} \times 1.5\ \mu\text{m}$ Permalloy element 50 nm thick was recorded at the L_3 -absorption edge of Ni (852.7 eV). The yellow arrows show the direction of the magnetization in the four domains, indicating the sense of flux closure of the Landau structure.

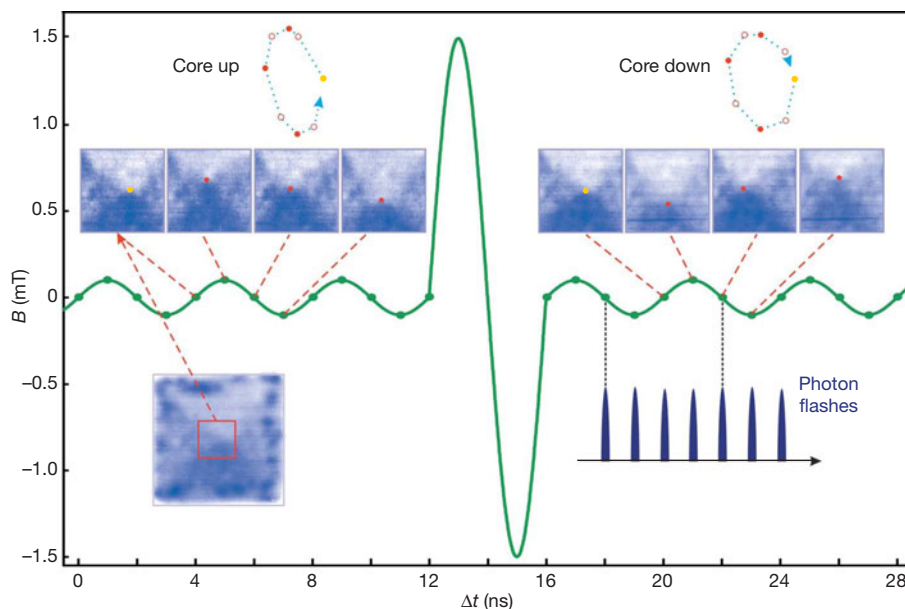


Figure 3 | Experimental results of switching the vortex core polarization. The vortex structure is excited with an alternating magnetic field (frequency 250 MHz, amplitude 0.1 mT). Two sequences (phase steps 90°) of images show the dynamic response of the vortex structure before and after a 4 ns 'single period' burst (amplitude 1.5 mT). Yellow dots indicate the vortex

core position at '0°', and red dots the position at the other phases. The vortex core positions extracted from the individual images form an elliptical trajectory. Unfilled red circles represent the vortex core position at the intermediate phase steps (see Supplementary Videos 1 and 2).

crosstalk between neighbouring patterns, all of which are indispensable features for ultra-high-density integration.

Thus, we have discovered a new switching scheme to change the direction of the vortex core polarization. Instead of fields of the order of 0.5 T, short in-plane bursts of 4 ns with an amplitude of only 1.5 mT were sufficient to toggle the vortex core polarization controllably. This phenomenon can be understood in the framework of micromagnetic theory and gives insight into the basic dynamics of magnetic nanostructures; it may open new perspectives for the application of magnetic vortex structures.

METHODS

The time- and space-resolved response of the sample magnetization was imaged by a stroboscopic measurement technique¹² with a scanning transmission X-ray microscope at the Advanced Light Source (beamline 11.0.2)²⁵. The contrast mechanism used for the imaging of the magnetic structures is the X-ray magnetic circular dichroism (XMCD) effect, namely the dependence of the absorption coefficient of circularly polarized X-rays on the magnetization of a ferromagnetic sample²⁰. The elliptically polarized monochromatic X-rays from the undulator

beamline are focused by a Fresnel-zone plate to a spot of about 30 nm. The sample is scanned through this spot with a high-resolution scanning stage under interferometric control. Images were recorded at the Ni L₃-absorption edge (852.7 eV), at which XMCD²⁰ gives a high magnetic contrast (the change in the absorption of circularly polarized photons is proportional to the projection of the magnetization on the photon propagation direction). To allow the observation of the in-plane Landau structure, the sample was placed at 60° to the incoming photon beam (Fig. 2). The sample was excited near its eigenfrequency with a 250-MHz RF signal and the response was monitored by varying the delay of the excitation signal between consecutive images. A time resolution of less than 100 ps is given by the inherent time structure of the synchrotron radiation. More details can be found in ref. 26.

The Permalloy (Ni₈₀Fe₂₀) samples and stripline structure were grown on a thin Si₃N₄ membrane (100 nm in thickness) with a transmission of 80% for photon energies of about 800 eV. They were patterned by e-beam lithography on to a Cu stripline 10 μm wide and 150 nm thick. The whole structure was capped with a 2-nm Al protective coating. The stripline generates a magnetic field in the plane of the sample perpendicular to the current direction.

A fast electronic pulse generator in combination with a broadband mixer was used to modulate the RF power to the sample. When no pulses are applied to the mixer, a small fraction of the input power reaches the sample and generates an alternating magnetic field with an amplitude of 0.1 mT. This is sufficient to induce a stable gyrotropic motion, observable with the microscope, and to deduce the sense of the vortex motion without inducing any vortex core switching. Superimposed on this background field, a single burst with amplitudes between 0.5 and 3 mT and lengths between 4 and 200 ns could be delivered to the sample. The phase relation between the burst and the RF signal can be adjusted by synchronizing the pulse generator with the RF signal.

The micromagnetic simulations were performed with OOMMF code²⁷. The time-dependent behaviour of a 1.5 μm × 1.5 μm × 50 nm Permalloy element was simulated with a cell size of 5 nm × 5 nm × 50 nm for Supplementary Video 3, and a cell size of 2 nm × 2 nm × 50 nm for Supplementary Video 4 (a two-dimensional lattice). Standard Permalloy material parameters were used: saturation magnetization $M_s = 860 \text{ kA m}^{-1}$, exchange constant $A = 13 \text{ pJ m}^{-1}$ and Landau-Lifshitz gyromagnetic ratio $\gamma = 2.21 \times 10^5 \text{ m A}^{-1} \text{ s}^{-1}$. The damping parameter α was set to 0.05. A linear in-plane alternating-current magnetic field was applied with an amplitude of 0.1 mT and a frequency of 250 MHz. The alternating field was followed by a short burst with a length of one monocycle and an amplitude of 3.5 mT. The time structure of the applied field was similar to that applied during the experiments (Fig. 3). Supplementary Video 4 is a zoomed-in movie of an area 150 nm × 150 nm showing only the switching process (720 ps).

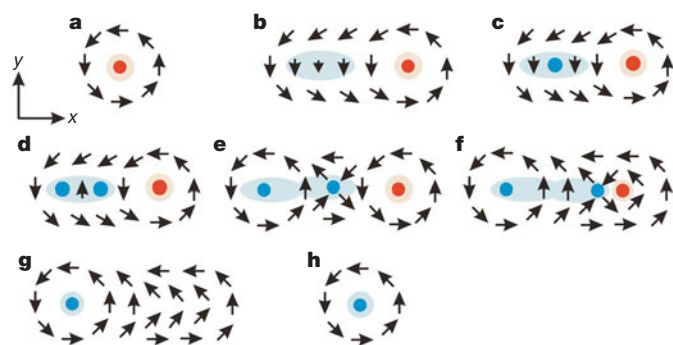


Figure 4 | Schematic representation of different steps in switching the vortex core polarization. The same conventions are used as for the two-dimensional schemes in Fig. 1. The arrows represent the in-plane magnetization components, and the coloured dot at the centre and the coloured rings around represent the out-of-plane component (blue, up; red, down). A detailed description is given in the text.

Received 19 October 2005; accepted 8 September 2006.

1. Hubert, A. & Schäfer, R. *Magnetic Domains—the Analysis of Magnetic Microstructures* (Springer, Berlin, 1998).
2. Raabe, J. *et al.* Magnetization pattern of ferromagnetic nanodisks. *J. Appl. Phys.* **88**, 4437–4439 (2000).
3. Shinjo, T., Okuno, T., Hassdorf, R., Shigeto, K. & Ono, T. Magnetic vortex core observation in circular dots of permalloy. *Science* **289**, 930–932 (2000).
4. Ross, C. *et al.* Micromagnetic behavior of electrodeposited cylinder arrays. *Phys. Rev. B* **65**, 144417 (2002).
5. Feldtkeller, E. & Thomas, H. Struktur und Energie von Blochlinien in dünnen ferromagnetischen Schichten. *Phys. Kondens. Mater.* **4**, 8–14 (1965).
6. Miltat, J. & Thiaville, A. Vortex cores—smaller than small. *Science* **298**, 555 (2002).
7. Wachowiak, A. *et al.* Direct observation of internal spin structure of magnetic vortex cores. *Science* **298**, 577–580 (2002).
8. Thiele, A. Steady-state motion of magnetic domains. *Phys. Rev. Lett.* **30**, 230–233 (1973).
9. Huber, D. Equation of motion of a spin vortex in a two-dimensional planar magnet. *J. Appl. Phys.* **53**, 1899–1900 (1982).
10. Choe, S. *et al.* Vortex core-driven magnetization dynamics. *Science* **304**, 420–422 (2004).
11. Park, J. & Crowell, P. Interactions of spin waves with a magnetic vortex. *Phys. Rev. Lett.* **95**, 167201 (2005).
12. Stoll, H. *et al.* High-resolution imaging of fast magnetization dynamics in magnetic nanostructures. *Appl. Phys. Lett.* **84**, 3328–3330 (2004).
13. Höllinger, R., Killinger, A. & Krey, U. Statics and fast dynamics of nanomagnets with vortex structure. *J. Magn. Magn. Mater.* **261**, 178–189 (2003).
14. Vavassori, P., Grimsditch, M., Metlushko, V., Zaluzec, N. & Ilic, B. Magnetoresistance of single magnetic vortices. *Appl. Phys. Lett.* **86**, 072507 (2005).
15. Taniuchi, T., Oshima, M., Akinaga, H. & Ono, K. Vortex-chirality control in mesoscopic disk magnets observed by photoelectron emission microscopy. *J. Appl. Phys.* **97**, 10J904 (2005).
16. Okuno, T., Shigeto, K., Ono, T., Mibu, K. & Shinjo, T. MFM study of magnetic vortex cores in circular permalloy dots: Behavior in external field. *J. Magn. Magn. Mater.* **240**, 1–6 (2002).
17. Thiaville, A., Garcia, J., Dittrich, R., Miltat, J. & Schrefl, T. Micromagnetic study of Bloch-point-mediated vortex core reversal. *Phys. Rev. B* **67**, 094410 (2003).
18. Argyle, B., Terrenzio, E. & Slonczewski, J. Magnetic vortex dynamics using the optical Cotton-Mouton effect. *Phys. Rev. Lett.* **53**, 190–193 (1984).
19. Park, J., Eames, P., Engebretson, D., Berezovsky, J. & Crowell, P. Imaging of spin dynamics in closure domain and vortex structures. *Phys. Rev. B* **67**, 020403 (2003).
20. Schütz, G. *et al.* Absorption of circularly polarized x-rays in iron. *Phys. Rev. Lett.* **58**, 737–740 (1987).
21. Novosad, V. *et al.* Magnetic vortex resonance in patterned ferromagnetic dots. *Phys. Rev. B* **72**, 024455 (2005).
22. Lee, K. S., Choi, S. & Kim, S. K. Radiation of spin waves from magnetic vortex cores by their dynamic motion and annihilation processes. *Appl. Phys. Lett.* **87**, 192502 (2005).
23. Döring, W. Point singularities in micromagnetism. *J. Appl. Phys.* **39**, 1006–1007 (1964).
24. Guslienko, K. *et al.* Eigenfrequencies of vortex state excitations in magnetic submicron-size disks. *J. Appl. Phys.* **91**, 8037–8039 (2002).
25. Kilcoyne, A. *et al.* Interferometer-controlled scanning transmission x-ray microscopes at the advanced light source. *J. Synchrotron Radiat.* **10**, 125–136 (2003).
26. Puzic, A. *et al.* Spatially resolved ferromagnetic resonance: Imaging of ferromagnetic eigenmodes. *J. Appl. Phys.* **97**, 10E704 (2005).
27. Donahue, M. & Porter, D. *OOMMF User's Guide, version 1.0. Interagency Report NISTIR 6376* (National Institute of Standards and Technology, Gaithersburg, Maryland, 1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. D. Carstanjen, D. Goll, S. Komineas, H. Kronmüller and M. Scheinfein for helpful discussions. Financial support was provided by the Deutsche Forschungsgemeinschaft through the priority programme 'Ultrafast Magnetisation Processes'. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.S. (stoll@mf.mpg.de).

Superconductivity in doped cubic silicon

E. Bustarret^{1*}, C. Marcenat^{3*}, P. Achatz^{1,3}, J. Kačmarčík^{1,4}, F. Lévy³, A. Huxley^{3†}, L. Ortéga², E. Bourgeois⁵, X. Blase⁵, D. Débarre⁶ & J. Boulmer⁶

Although the local resistivity of semiconducting silicon in its standard crystalline form can be changed by many orders of magnitude by doping with elements, superconductivity has so far never been achieved. Hybrid devices combining silicon's semiconducting properties and superconductivity have therefore remained largely underdeveloped. Here we report that superconductivity can be induced when boron is locally introduced into silicon at concentrations above its equilibrium solubility. For sufficiently high boron doping (typically 100 p.p.m.) silicon becomes metallic¹. We find that at a higher boron concentration of several per cent, achieved by gas immersion laser doping, silicon becomes superconducting. Electrical resistivity and magnetic susceptibility measurements show that boron-doped silicon (Si:B) made in this way is a superconductor below a transition temperature $T_c \approx 0.35$ K, with a critical field of about 0.4 T. *Ab initio* calculations, corroborated by Raman measurements, strongly suggest that doping is substitutional. The calculated electron–phonon coupling strength is found to be consistent with a conventional phonon-mediated coupling mechanism². Our findings will facilitate the fabrication of new silicon-based superconducting nanostructures and mesoscopic devices with high-quality interfaces.

Semiconductors, such as GeTe, PbTe and SrTiO₃, have been shown in the past to become superconducting at low temperature upon carrier doping^{3–5}. In contrast, superconductivity has been sought in silicon for 60 years⁶ but observed only in its β -Sn and hexagonal (sh) metallic phases obtained under extreme pressures^{7–9} of the order of 10 GPa. This is surprising, given that theoretical calculations have consistently predicted doped germanium¹⁰ and silicon^{11,12} to be superconducting in their face-centred cubic (f.c.c.; ambient pressure) diamond structure and that conventional superconductivity has been found in materials that contain light group IV elements, such as silicon clathrates^{11,13} and boron-doped diamond^{14–16}.

Here we report structural, electrical and magnetic experiments performed at room pressure on a set of boron-doped silicon thin layers where boron was incorporated into cubic silicon well above its equilibrium solubility (1.2 at.% or $6 \times 10^{20} \text{ cm}^{-3}$). As depicted in Fig. 1, this was achieved by gas immersion laser doping (GILD), where the precursor gas (BCl₃) is chemisorbed on a (001)-oriented silicon wafer before each ultraviolet laser pulse^{17–19}. During each melting/solidification cycle, boron diffuses into molten silicon, and is incorporated at substitutional sites of the crystal as the liquid/solid interface moves back to the surface of the epilayer¹⁷. The in-plane lattice matching to the substrate and the local incorporation of boron atoms of lower covalent radius²⁰ result in the formation of biaxially strained pseudomorphic layers^{17,19} with thicknesses ranging from 10 nm to 120 nm.

High-resolution X-ray diffraction (XRD) analysis of three samples was performed around the (004) Bragg reflection. As shown in Fig. 1,

we observed one or two broad diffraction peaks well separated from the narrow line from the substrate. This indicates a non-uniform distribution of the strain perpendicular to the surface and thus of the boron content, but confirms the epitaxial single crystal character of the films. The two broad maxima detected in sample 1 at 35.56° and 36.02° correspond to a contraction of the lattice parameter, a , along (001) of $da/a = -2.5\%$ and -3.7% , respectively. Assuming that the Si:B alloy retains the elastic properties of bulk silicon, this yields an in-plane tensile stress ('negative' pressure) of 2.6 GPa and 3.9 GPa, and isotropic lattice parameter variations of -1.4% and -2.1% . From Vegard's law and previous calibrations^{17,19,20}, this corresponds to substitutional boron concentrations of respectively 5.7 at.% and 8.4 at.% (that is, 2.8×10^{21} and $4.2 \times 10^{21} \text{ B cm}^{-3}$). Hall measurements performed at room temperature on the same sample (sample 1) give a free-hole density of the order of

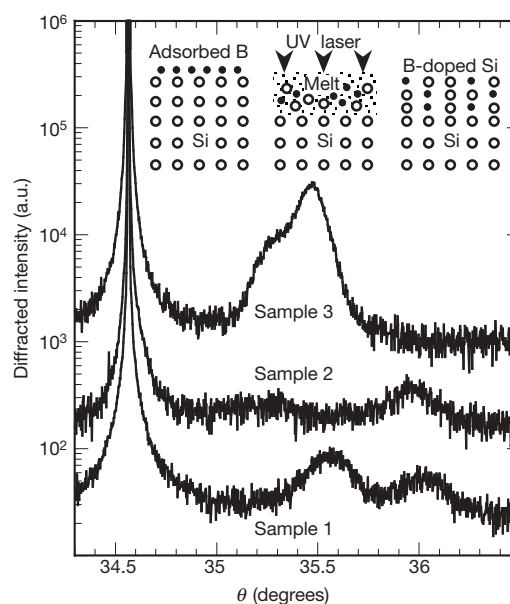


Figure 1 | High-resolution XRD measurements of three Si:B superconducting samples, and diagrams showing three steps of the GILD process. The diffraction peaks have the following Bragg angles for the (004) reflection: 34.565° (Si substrate); 35.56° and 36.02° (sample 1, top curve); 35.23° and 35.97° (sample 2, middle curve); and 35.30° and 35.47° (sample 3, bottom curve). Inset, diagrams of the three initial steps of gas immersion laser doping (GILD). These steps, repeated about 200 times, are: adsorption of BCl₃ on the silicon wafer surface (left); local surface melting of silicon induced by the laser shot (centre); and formation of the Si:B epilayer upon cooling (right).

¹Laboratoire d'Etudes des Propriétés Electroniques des Solides, ²Laboratoire de Cristallographie, CNRS, BP166, 38042 Grenoble, France. ³Département de la Recherche Fondamentale sur la Matière Condensée, SPSMS, CEA Grenoble, 17 rue des Martyrs, 38054 Grenoble, France. ⁴Institute of Experimental Physics, Slovak Academy of Sciences, SK-04001 Košice, Slovakia. ⁵Laboratoire de Physique de la Matière Condensée et Nanostructures, Université Lyon I and CNRS, Domaine scientifique de la Doua, 69622 Villeurbanne, France. ⁶Institut d'Electronique Fondamentale, Université Paris Sud and CNRS, Bât. 220, 91405 Orsay, France. [†]Present address: School of Physics, University of Edinburgh, Edinburgh EH9 3JZ, UK.

*These authors contributed equally to this work.

$(5 \pm 2) \times 10^{21} \text{ cm}^{-3}$, which is comparable to the higher boron concentration deduced above. The XRD results shown in Fig. 1 indicate that sample 2 has a slightly lower B concentration, whereas sample 3 is thicker and has a much lower concentration than sample 1.

We now discuss the electrical characterization of the samples. The room-temperature sheet resistances (in Ω per $\square$) of samples 1, 2 and 3 are 35, 17 and 8, respectively, while their resistivities are of the order of $100 \mu\Omega \text{ cm}$, and have a dependence on temperature over the 0.35–300 K range typical of disordered metals (see Fig. 2). The inset of Fig. 2 shows measurements made to lower temperature with a dilution refrigerator. Two of the samples (1 and 2) undergo superconducting transitions to zero resistance, respectively at 150 mK and 45 mK, whereas sample 3 shows only a partial transition. It should be noted that elemental boron becomes superconducting²¹ only under a very high pressure (of the order of 160 GPa) and therefore cannot account for this superconductivity.

The temperature dependence below 0.5 K of the d.c. electrical resistance, R , and of the real part of the a.c. magnetic susceptibility, χ' , for sample 1 is plotted in Fig. 3a. A sharp drop of the resistance is observed with an onset around 0.4 K. An immeasurably small R value is observed below 150 mK. Similarly, as the temperature decreases there is an onset of diamagnetism below 0.34 K, but full magnetic screening is not achieved until 150 mK. These measurements unambiguously demonstrate the occurrence of superconductivity in sample 1. The foot of the resistive and diamagnetic transitions, and the width of the magnetic response, are typical of an inhomogeneous superconductor²²; this inhomogeneity is expected, given the non-uniform boron distribution detected by XRD. In order to define the phase diagram in the magnetic field–temperature (H – T) plane, we performed electrical measurements; these involved varying the temperature at different magnetic fields, and varying H at different temperatures (see the insets in Fig. 2b), with the field always applied perpendicular to the plane of the film. The transitions remain sharp and no hysteresis or supercooling was seen, which is consistent with the transition to superconductivity under field being of second order.

This observation, and the relatively high value of the critical field, 0.4 T (see Fig. 3b), which exceeds that of lead ($8 \times 10^{-2} \text{ T}$), the strongest known type I superconductor, suggests that boron-doped silicon is a type II superconductor. The critical temperature T_{c2} and the

critical magnetic field H_{c2} where the resistance is 10% of its normal state value are plotted in Fig. 3b (the 10% criterion corresponds to the onset of diamagnetism in zero field). To indicate the width of the transition, the positions of maxima of the slope of the resistance versus temperature (dR/dT) in the T -sweeps and maxima of dR/dH in the H -sweeps are also shown (open symbols in Fig. 3b). The critical field versus temperature curve in Fig. 3b has a marked curvature compared with the standard Bardeen–Cooper–Schrieffer (BCS) dependence²³. This feature could be explained by paramagnetic limitation if the carriers are assigned a gyromagnetic factor, g , 40% larger²⁴ than the standard value of 2. Alternatively, the observed dependence might be a consequence of the film's inhomogeneity: it could arise if the dimensions of the superconducting regions are restricted parallel to the film surface, or be due to percolation or localization phenomena^{25,26}. The Landau–Ginzburg superconducting coherence length is estimated in the orbital limit to be 13 nm from the experimentally measured slope of H_{c2} against temperature at T_{c2} , whereas a larger estimate of approximately 20 nm is found based on the values of T_c and $H_{c2}(0)$ (at zero temperature) within the standard BCS theory. These values are not much smaller than the estimated thickness of the film ($35 \pm 5 \text{ nm}$), which can also act to limit the extent of the coherence length²⁷.

In order to provide some insight into the microscopic pairing mechanism, we performed an *ab initio* study of the electronic, vibrational and electron–phonon coupling properties of boron-doped silicon. We adopted a supercell approach, with one boron atom

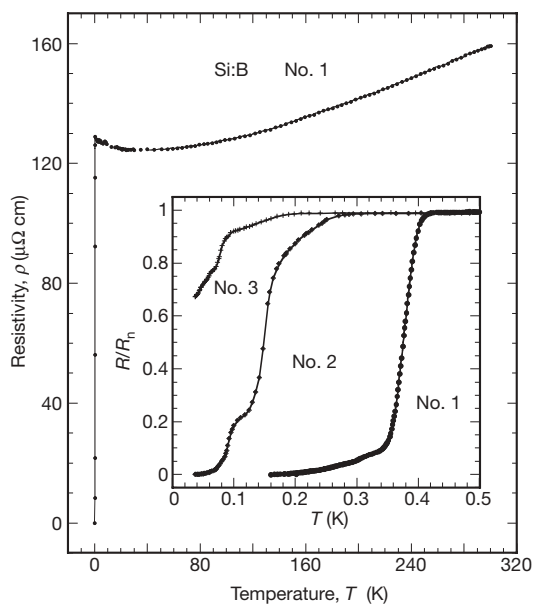


Figure 2 | Temperature dependence of the a.c. resistivity, ρ . Main figure, the metallic behaviour of crystalline silicon doped with boron in the at.% range (sample 1). Inset, temperature dependence of the a.c. resistance at low temperatures for samples 1, 2 and 3, normalized to their normal state resistance, R_n , at 2 K.

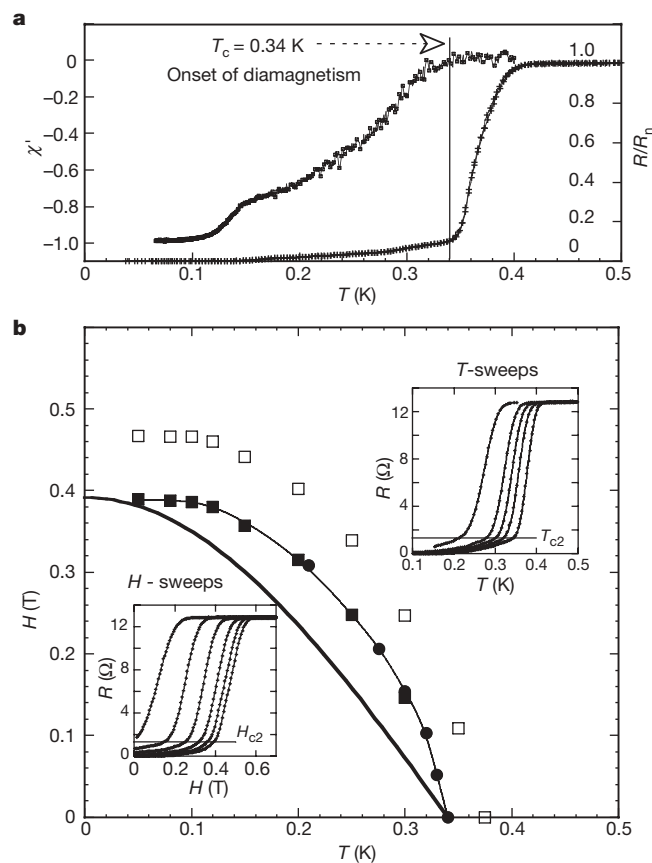


Figure 3 | Superconducting transition and magnetic phase diagram of B-doped Si (sample 1). **a**, Temperature dependence of the d.c. electrical resistance R (normalized to its normal state value R_n) and of the real part of the a.c. susceptibility. **b**, Temperature dependence of the upper critical field H_{c2} . Filled squares, H_{c2} measured by field sweeps at 50, 150, 200, 250, 300 and 350 mK (see lower inset); filled circles, T_{c2} measured by temperature sweeps at 0.1, 0.2, 0.3, 0.4 and 0.5 T (see upper inset). Open symbols, location of dR/dT and dR/dH maxima for the curves reported in the insets; full line, classical theory¹⁹.

substituted for every 16 silicon atoms arranged in an f.c.c. cell, yielding a 6.25 at.% concentration, close to the values estimated for the superconducting samples (see Methods). Structural relaxation led to an isotropic contraction of 1.9% of the lattice parameter compared to the pure silicon, related to the reduced Si–B bond length (2.1 Å in our calculations), and within the range of values deduced experimentally by XRD.

Close to the Fermi level (which is located 0.5 eV below the top of the valence bands), the calculated band structure resembles that of undoped silicon, confirming the degenerate nature of Si:B at such a high doping level. The phonon density of states (pDOS) is represented in Fig. 4. As compared to undoped silicon, we found a $\sim 42\text{ cm}^{-1}$ softening of the zone-centre optical modes. The softening is caused by hole injection, which more than compensates the effect of lattice contraction. As illustrated in Fig. 4, the calculated softening agrees well with the broadened Raman peaks observed in samples 1, 2 and 3 at 464, 480 and 485 cm^{-1} respectively, and can be compared with a peak position of 520 cm^{-1} in pure Si. Experimentally, the shift to lower frequency and the broadening are more pronounced for sample 1, which has the highest T_c . The appearance of an additional Raman-active Si–B stretching mode 100 cm^{-1} above the zone-centre optical modes, seen experimentally, is found in the calculation to correspond to a Si–B stretching mode at around 590–600 cm^{-1} . This observation further confirms the substitutional incorporation of boron.

To estimate the superconducting transition temperature, we calculated the electron–phonon coupling function (Eliashberg function), which measures the coupling strength as a function of the phonon energy (Fig. 4). As for boron-doped diamond^{12,28,29}, most of the coupling originates from the optical modes. Averaging over the Brillouin zone and phonon bands, we find a coupling constant λ of 0.28. This is in good agreement with $\lambda = 0.30$ for 5% doping within the virtual crystal approximation¹². The peaked structure of the Eliashberg function around $\omega_0 = 470\text{ cm}^{-1}$ suggests the use of the following modified McMillan formula appropriate for a δ -shaped spectrum: $T_c = \hbar\omega_0 \exp[-(1 + \lambda)/(1 - \mu^*(1 + \lambda))]$, where μ^* accounts for the screened Coulomb repulsion. We find that for μ^*

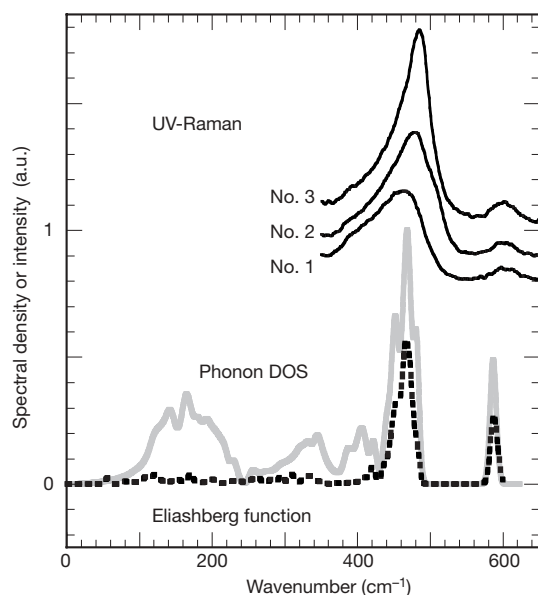


Figure 4 | Calculated and experimental vibrational spectral density of B:Si. Phonon density of states (thick grey line) and Eliashberg function (broken line) as calculated *ab initio* for a B:Si₁₅ supercell, compared to experimental UV-Raman spectra of samples 1, 2 and 3 in the 350–650 cm^{-1} range. For the sake of comparison with experiment, the theoretical frequencies have been rescaled by $520/502 = 1.036$. Please note that at the present excitation energy (3.82 eV), the absorption length in Si:B is so short that no signal from the substrate could be detected in all the samples.

in the range 0.08–0.12 (a value of $\mu^* = 0.1$ was adopted in ref. 12), T_c is predicted to vary in the range 0.5–0.03 K. This suggests that a standard BCS mechanism can account for the observed superconducting transition. Theory¹² as well as experimental results obtained on superconducting boron-doped diamond³⁰ lead us to expect that a higher superconducting transition temperature might be achieved if more boron could be incorporated into the silicon. This is an exciting possibility, as it would considerably facilitate the development of Si:B nanostructures and mesoscopic devices exploiting superconductivity.

METHODS

Sample preparation and characterization. Laser-induced doping was performed with a 308 nm wavelength laser over ten separate 6 mm² areas of the same wafer in a special GILD configuration. The precursor gas (BCl₃) was injected and chemisorbed on a {001}-oriented silicon wafer surface before each of the 200 laser shots of 25 ns duration. *In situ* monitoring of the transient reflectivity at 675 nm, as well as previous calibrations by secondary ion mass spectroscopy¹⁹, enabled us to adjust the power and duration of the pulses in order to ensure maximum doping over the whole thickness of the melt. Estimated thicknesses (in nm) for samples 1, 2 and 3 were 35 ± 5 , 30 ± 5 and 110 ± 10 , respectively. Micro-Raman backscattering spectra excited by a c.w. HeCd laser (3 mW at 325 nm) were collected at room temperature in air under a $\times 40$ microscope objective. We used a HR800-UV Jobin-Yvon-Horiba spectro-photometer equipped with an edge filter and a liquid-N₂-cooled $256 \times 2,048$ pixel CCD detector. The spectral resolution was better than 2 cm^{-1} . High-resolution XRD measurements were performed with a commercial Philips Materials research diffractometer at the Cu K α_1 wavelength of 0.15406 nm selected with a (220)Ge four-reflection Bartels-type monochromator. Rocking curves were obtained by measuring the diffracted intensity as a function of the rotation, θ , of the sample. Data were collected around the (004) silicon Bragg reflection with an angular resolution better than 3×10^{-3} degree. The absence of polycrystalline phases was checked by additional diffraction measurements at grazing incidence.

Electrical and magnetic measurements. Electrical resistivity measurements were performed from room temperature down to 0.35 K using a Quantum Design Physical Properties Measurements System and were extended down to 40 mK in dilution refrigerators using standard lock-in techniques. Gold wires were attached onto the samples in a four-terminal configuration either directly with silver epoxy or on evaporated Au:Ti contacts for a better definition of geometrical factors. To avoid artefacts, various careful checks were made, such as the comparison of a.c. and d.c. electrical measurements as well as current–voltage I – V characteristics. The normalized resistance values displayed in Fig. 3a result from d.c. measurements with $I = 100$ nA.

Measurements of a.c. magnetic susceptibility were obtained by recording the change in the mutual inductance of a small coil at 9 kHz with a Stanford Research 836 lock-in amplifier. The coil response was calibrated by measuring a thin film of niobium with similar size and shape. All magnetic measurements were performed with the magnetic field perpendicular to the film plane.

Ab initio calculations. Our calculations were based on an ultrasoft pseudo-potential plane-wave implementation of density functional theory with a generalized-gradient description of the exchange–correlation functional. They were performed with the Plane Wave Self-Consistent Field package (www.pwscf.org). The Perdew–Burke–Ernzerhof functional was used, together with a 25 Ryd and a 200 Ryd cut-off for the eigenstates and charge density, respectively. Electronic and phonon momentum were sampled with a $5 \times 5 \times 5$ and $4 \times 4 \times 4$ Monkhorst–Pack grid, respectively. The electronic sampling was increased to $10 \times 10 \times 10$ for evaluating λ . Within the present approach, the zone centre optical modes for pure cubic silicon are found at 502 cm^{-1} , compared to the experimental value of 520 cm^{-1} . For the sake of comparison with experiment, the theoretical frequencies have been rescaled by $520/502 = 1.036$ in Fig. 4.

Received 29 June; accepted 9 October 2006.

1. Dai, P., Zhang, Y. & Sarachik, M. P. Critical conductivity exponent for Si:B. *Phys. Rev. Lett.* **66**, 1914–1917 (1991).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
3. Hein, R. A., Gibson, J. W., Mazelsky, R., Miller, R. C. & Hulm, J. K. Superconductivity in germanium telluride. *Phys. Rev. Lett.* **12**, 320–322 (1964).
4. Schooley, J. F. *et al.* Dependence of the superconducting transition temperature on carrier concentration in semiconducting SrTiO₃. *Phys. Rev. Lett.* **14**, 305–307 (1965).
5. Lasbley, A., Granger, R. & Rolland, S. High temperature superconducting behaviour in PbTe–Pb system. *Solid State Commun.* **13**, 1045–1048 (1973).

6. Alekseyevsky, N. & Migunov, L. Investigation of metals at temperatures below 1°K. *J. Phys. (USSR)* **11**, 95 (1947).
7. Buckel, W. & Wittig, J. Supraleitung von Germanium und Silizium unter hohem Druck. *Phys. Lett.* **17**, 187–188 (1965).
8. Stepanov, G. N., Valyanskaya, T. V. & Yakovlev, E. N. Superconductivity of metallic silicon below the pressure of transition to metallic modification. *Sov. Phys. Solid State* **22**, 292–293 (1980).
9. Chang, K. J. *et al.* Superconductivity in high-pressure metallic phases of Si. *Phys. Rev. Lett.* **54**, 2375–2378 (1985).
10. Cohen, M. L. Superconductivity in many-valley semiconductors and in semimetals. *Phys. Rev.* **134**, A511–A521 (1964).
11. Connétable, D. *et al.* Superconductivity in doped sp^3 semiconductors: The case of the clathrates. *Phys. Rev. Lett.* **91**, 247001 (2003).
12. Boeri, L., Kortus, J. & Andersen, O. K. Three-dimensional MgB_2 -type superconductivity in hole-doped diamond. *Phys. Rev. Lett.* **93**, 237002 (2004).
13. Kawaji, H., Horie, H., Yamanaka, S. & Ishikawa, M. Superconductivity in the silicon clathrate compound $(Na,Ba)_xSi_{46}$. *Phys. Rev. Lett.* **74**, 1427–1429 (1995).
14. Ekimov, E. A. *et al.* Superconductivity in diamond. *Nature* **428**, 542–545 (2004).
15. Yokoya, T. *et al.* Origin of the metallic properties of heavily boron-doped superconducting diamond. *Nature* **438**, 647–650 (2005).
16. Sacépé, B. *et al.* Tunneling spectroscopy and vortex imaging in boron-doped diamond. *Phys. Rev. Lett.* **96**, 097006 (2006).
17. Kerrien, G. *et al.* Ultra-shallow, super-doped and box-like junctions realized by laser-induced doping. *Appl. Surf. Sci.* **186**, 45–51 (2002).
18. Kerrien, G. *et al.* Gas immersion laser doping (GILD) for ultra-shallow junction formation. *Thin Solid Films* **453–454**, 106–109 (2004).
19. Kerrien, G. *et al.* Optical characterization of laser-processed ultrashallow junctions. *Appl. Surf. Sci.* **208–209**, 277–284 (2003).
20. Vailionis, A., Glass, G., Desjardins, P., Cahill, D. G. & Greene, J. E. Electrically active and inactive B lattice sites in ultrahighly B doped Si(001): An X-ray near-edge absorption fine-structure and high-resolution diffraction study. *Phys. Rev. Lett.* **82**, 4464–4467 (1999).
21. Eremets, M. I., Struzhkin, V. V., Mao, H.-k. & Hemley, R. J. Superconductivity in boron. *Science* **293**, 272–274 (2001).
22. Tinkham, M. in *Inhomogeneous Superconductors-1979* (eds Gubser, D. U., Francavilla, T. L., Laibowitz, J. R. & Wolf, S. A.) 1–12 (AIP Conf. Proc. Vol. 58, American Institute of Physics, 1980).
23. Werthamer, N. R., Helfand, E. & Hohenberg, P. C. Temperature and purity dependence of the superconducting critical field H_{c2} , III. Electron spin and spin-orbit effects. *Phys. Rev.* **147**, 295–302 (1966).
24. Ting, C. S., Lee, T. K. & Quinn, J. J. Effective mass and g factor of interacting electrons in the surface inversion layer of silicon. *Phys. Rev. Lett.* **34**, 870–874 (1975).
25. Kent, A. D., Kapitulnik, A. & Geballe, T. H. H_{c2} spectroscopy of geometrical effects in La-S Films. *Phys. Rev. B* **36**, 8827–8830 (1987).
26. Quateman, J. H. T_c suppression and critical fields in thin superconducting Nb films. *Phys. Rev. B* **34**, 1948–1951 (1986).
27. Guyon, E., Meunier, F. & Thomson, R. S. Thickness dependence of κ_2 and related problems for superconducting alloy films in strong fields. *Phys. Rev.* **156**, 452–469 (1967).
28. Blase, X., Adessi, Ch & Connétable, D. Role of the dopant in the superconductivity of diamond. *Phys. Rev. Lett.* **93**, 237004 (2004).
29. Lee, K.-W. & Pickett, W. E. Superconductivity in boron-doped diamond. *Phys. Rev. Lett.* **93**, 237003 (2004).
30. Bustarret, E. *et al.* Dependence of the superconducting transition temperature on the doping level in single-crystalline diamond films. *Phys. Rev. Lett.* **93**, 237005 (2004).

Acknowledgements The authors acknowledge J. Marcus, M. Sanquer and X. Jehl for access to their cryostats, as well as J. Pernot for discussions. Calculations were performed at the CNRS national supercomputing centre (IDRIS). Partial funding by the French ANR-05-BLAN programme is acknowledged.

Author Contributions The samples were prepared by D.D. and J.B., and the calculations performed by E.Bo. and X.B. All other authors contributed to the physical characterization of the samples.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E. Bu. (etienne.bustarret@grenoble.cnrs.fr) and C.M. (christophe.marcenat@cea.fr).

Watching hydrogen-bond dynamics in a β -turn by transient two-dimensional infrared spectroscopy

Christoph Kolano¹, Jan Helbing¹, Mariusz Kozinski¹, Wolfram Sander² & Peter Hamm¹

X-ray crystallography and nuclear magnetic resonance measurements provide us with atomically resolved structures of an ever-growing number of biomolecules. These static structural snapshots are important to our understanding of biomolecular function, but real biomolecules are dynamic entities that often exploit conformational changes and transient molecular interactions to perform their tasks. Nuclear magnetic resonance methods can follow such structural changes, but only on millisecond timescales under non-equilibrium conditions. Time-resolved X-ray crystallography has recently been used to monitor the photodissociation of CO from myoglobin on a subnanosecond timescale¹, yet remains challenging to apply more widely. In contrast, two-dimensional infrared spectroscopy, which maps vibrational coupling between molecular groups and hence their relative positions and orientations^{2–11}, is now routinely used to study equilibrium processes on picosecond timescales. Here we show that the extension of this method into the non-equilibrium regime^{12,13} allows us to observe in real time in a short peptide the weakening of an intramolecular hydrogen bond and concomitant opening of a β -turn. We find that the rate of this process is two orders of mag-

nitude faster than the ‘folding speed limit’ established for contact formation between protein side chains¹⁴.

Our two-dimensional infrared (2D-IR) investigation of ultrafast dynamics uses the cyclic disulphide-bridged peptide cyclo(Boc-Cys-Pro-Aib-Cys-OMe) shown in Fig. 1a (for details regarding synthesis and characterization, see ref. 15). The disulphide bridge and an intramolecular hydrogen bond make the peptide very rigid and ensure that it adopts one exclusive β -turn structure, according to the nuclear magnetic resonance (NMR) results¹⁵. The relatively weak disulphide bridge (dissociation energy ~ 65 kcal mol⁻¹) is easily cleaved by ultraviolet light and can thus serve as a photo-trigger for the creation of a predetermined breaking point^{16,17}. The Fourier transform infrared absorption spectrum of the peptide shows well-resolved bands in the amide I region, which are associated with individual C=O groups of the peptide backbone (Fig. 1b; see Supplementary Information for band assignment).

In a first control experiment, we use conventional transient (pump-probe) infrared spectroscopy on a 200 mM solution of the peptide in CD₃CN to establish the dynamic processes in our system and their timescales (see Supplementary Information for full

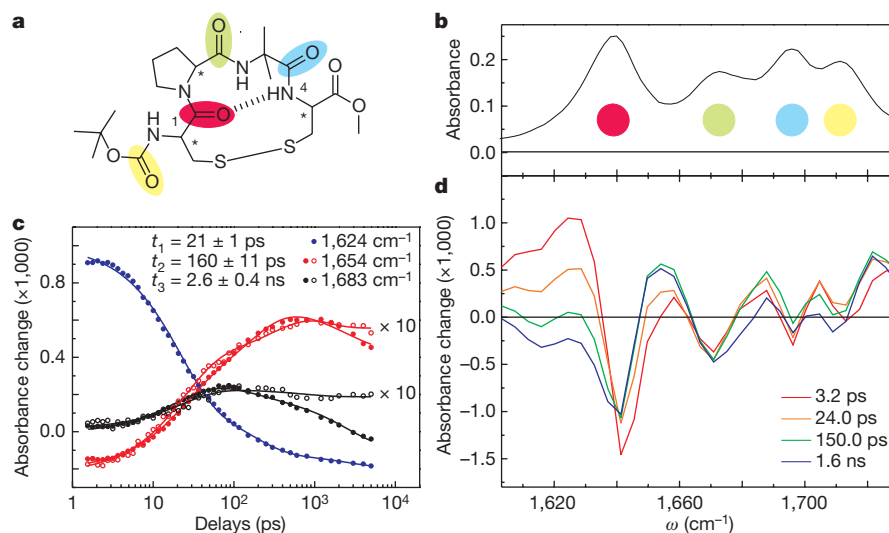


Figure 1 | Absorption and transient 1D spectra. **a**, Chemical structure of the cyclic disulphide-bridged tetrapeptide cyclo(Boc-CPUC-OMe). The dashed line indicates the intramolecular hydrogen bond. **b**, Fourier transform infrared absorption spectrum in the amide I region of the peptide. Band assignments correspond to the coloured regions of the chemical structure of the peptide shown in **a**. **c**, Dynamics at selected spectral positions (filled circles, 200 mM; open circles, 14 mM). The solid lines correspond to a

simultaneous fit of the time traces by a sum of three exponentials. The data of the low concentration measurement are magnified by a factor of ten.

d, Difference infrared absorption spectra at different time delays after photolysis of the peptide with a ultraviolet laser flash. Bands appearing on irradiation are pointing upward; bands disappearing are pointing downward.

¹Physikalisch-Chemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH8057 Zürich, Switzerland. ²Lehrstuhl für Organische Chemie II, Ruhr-Universität Bochum, Universitätsstrasse 150, D44801 Bochum, Germany.

experimental methods). This reveals that the photolysis of the S–S bond results in an instantaneous redshift of all amide I bands (Fig. 1d, red curve), and that the subsequent evolution of the spectrum is characterized by three global time constants with values of ~ 20 ps, 160 ps and 2.6 ns (see Fig. 1c).

We assign the instantaneous redshift at early times to transient heating of the molecule by the excess energy that is released after cleavage of the S–S bond by the ultraviolet photon, a common effect in ultraviolet–infrared experiments^{18,19}. The heating leads to an excitation of low-frequency modes that are anharmonically coupled to the amide I vibrations²⁰. In accordance with this interpretation, the redshift disappears on a 20 ps timescale as the molecules cool through dissipation of excess energy into the solvent.

For pump–probe delay times longer than 1 ns, the measured signal depends on the peptide concentration used (compare open and closed symbols in Fig. 1c). This implies that the slowest dynamic process we probe (time constant of ~ 2.6 ns at 200 mM; closed circles) is the diffusion-controlled intermolecular reaction of the liberated thiyl radicals with other radicals or unreacted peptides; decreasing the peptide concentration makes the intramolecular radical recombination even slower (Fig. 1c, open circles). This behaviour contrasts with that seen in an earlier study of peptide conformational changes upon disulphide photocleavage¹⁶, where 90% of thiyl radicals recombined within 2–5 ns. Molecular dynamics simulations (see below) indicate that in our system, the strain in the much shorter peptide backbone keeps the two S radicals apart and prevents intramolecular recombination.

The intermediate process (~ 160 ps) is best seen in the absorption changes at $1,654\text{ cm}^{-1}$ (Fig. 1c), which reflect a blueshift of the C=O group of Cys(1) involved in the peptide's intramolecular hydrogen

bond. It is well known that in such a hydrogen-bond pattern, weakening of the hydrogen bond increases the bond order of the C=O bond and hence its vibrational frequency²¹. As a working hypothesis, we thus assign the second time constant to a conformational change of the peptide backbone, which is accompanied by a weakening of the intramolecular hydrogen bond.

In a second control experiment, we measure the conventional 2D-IR spectrum under equilibrium conditions, with the S–S bond still intact. To illustrate briefly the principle of 2D-IR spectroscopy, consider a system of two coupled carbonyl groups¹¹. Excitation of one carbonyl group results in a so-called diagonal absorbance signal at the frequency of the excited oscillator, and in a so-called off-diagonal absorbance signal (or crosspeak) at the frequency of the coupled second carbonyl group. Each of these signals consists of a negative contribution due to bleach and stimulated emission, and a positive contribution due to excited-state absorption. The latter is slightly redshifted owing to the anharmonicity of the carbonyl vibrators. Figure 2b displays the weighted differences between 2D-IR spectra recorded with parallel and perpendicular polarizations of pump and probe laser pulses, a procedure that enhances crosspeaks relative to the otherwise dominant diagonal peaks¹⁰. Most crosspeaks are caused by coupling of neighbouring amide groups, that is, between Cys(1)–Pro (labelled '1' in Fig. 2b) and Cys(1)–Boc (labelled '2' in Fig. 2b). Of importance will be the additional crosspeak that arises owing to the interaction between Cys(1) and Aib (labelled '3' in Fig. 2b). Because the peptide is clasped by the disulphide bridge and stabilized by an intramolecular hydrogen bond, the spatial distance between the amide groups of Cys(1) and Aib is sufficiently small to give rise to a crosspeak, albeit one that is hardly resolved because it is covered by the wings of the stronger crosspeaks labelled 1 and 2 in Fig. 2b.

With these results in mind, we turn to transient 2D-IR spectroscopy, which can be thought of as a combination of the two previous experiments. We initiate the experiment with an ultraviolet pulse that cleaves the disulphide bridge and then use increasingly longer delays before starting the 2D-IR part of the experiment, thus recording a 2D-IR spectrum as the molecule undergoes its photo-triggered conformational transition. Because the 2D-IR spectrum in the presence of the ultraviolet pulse contains contributions from the photo-product and from unexcited molecules, two sets of 2D-IR spectra are recorded simultaneously, one with the ultraviolet pulse switched on and one with the ultraviolet pulse switched off. Subsequent subtraction leads to transient 2D-IR difference spectra¹², which are sensitive to changes during the conformational transition. Transient 2D-IR spectroscopy thus extends 2D-IR spectroscopy, to allow the investigation of a transient species far from equilibrium with picosecond time resolution.

Transient 1D-IR spectra and transient 2D-IR spectra obtained with pump–probe delay times of 3 ps, 25 ps and 100 ps are shown in Fig. 3. The shift of all bands in the transient 1D-IR spectrum (Fig. 3a) gives rise to related signals along the diagonal in the transient 2D-IR spectra (Fig. 3b). But the most important feature is a transient crosspeak labelled 'TC' in Fig. 3b, which matches the position of crosspeak '3' (Cys(1)–Aib) in the equilibrium 2D-IR spectrum. The evolution of this crosspeak is seen more clearly in Fig. 3c, where the red curves show how absorbance changes with probe laser frequency for a fixed pump laser frequency. (The curves thus constitute horizontal cuts through the transient 2D-IR spectra—indicated by red dotted lines in Fig. 3b—at a pump frequency where the crosspeak occurs.) The absolute crosspeak intensity increases with delay time, and even more so relative to the intensity of the corresponding diagonal signal (Fig. 3c). Although the exact transient 2D-IR lineshape critically depends on multiple cancellation effects in the two-dimensional difference spectrum¹² that are at present inaccessible by theoretical modelling, the diagonal cut in Fig. 3c (blue line) does suggest the characteristic ' $\pm$ ' signature expected for vibrational transitions. We see a transient crosspeak only for Cys(1)–Aib, and not for Cys(1)–Pro or Cys(1)–Boc (although these latter crosspeaks, labelled

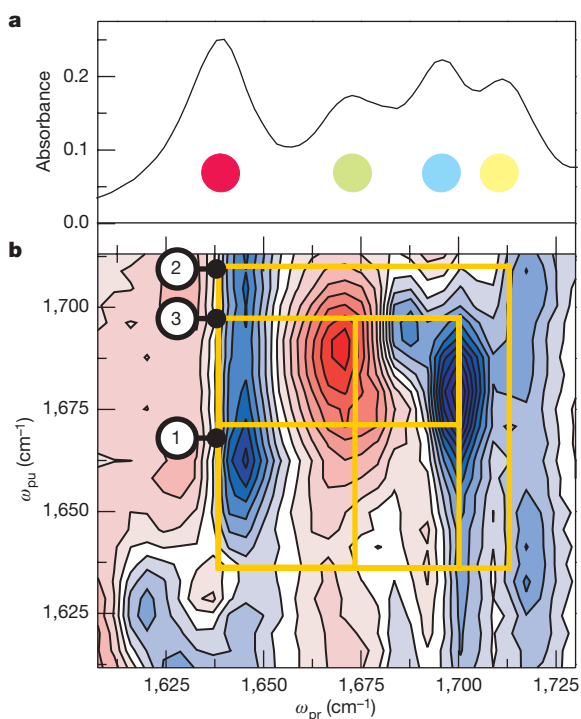


Figure 2 | Equilibrium 2D-IR spectra. **a**, Fourier transform infrared absorption spectrum in the amide I region of the peptide. Band assignments correspond to the coloured regions of the chemical structure of the peptide shown in Fig. 1a. **b**, Weighted difference between perpendicular and parallel polarization 2D-IR spectra. Negative signals are depicted in blue, positive signals in red. Crosspeaks due to coupling of neighbouring amide groups are labelled '1' and '2', whereas '3' highlights the crosspeak across the intramolecular hydrogen bond. ω_{pu} and ω_{pr} are the pump and probe frequencies in the 2D-IR experiment, respectively. The corners of the yellow squares link diagonal and corresponding crosspeaks.

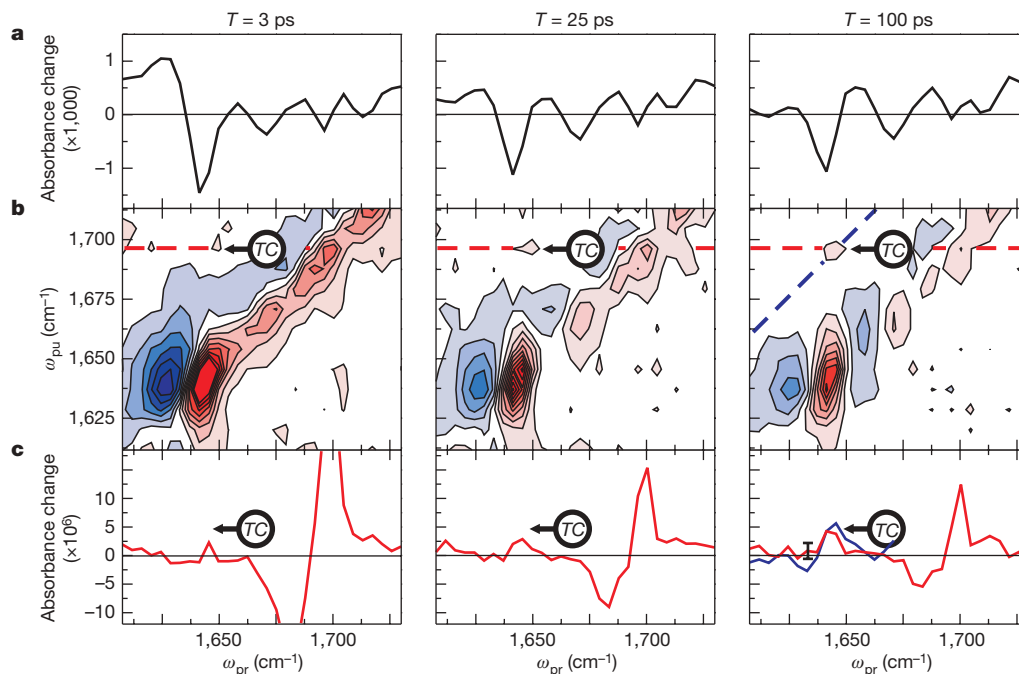


Figure 3 | Transient 2D-IR spectra. **a**, Reference transient 1D-IR spectra at delay times of 3 ps, 25 ps and 100 ps after photolysis. **b**, Transient 2D-IR spectra at ultraviolet pump to 2D-IR probe delay times of 3 ps, 25 ps and 100 ps with parallel polarization of all pulses. In the transient two-dimensional spectra, negative signals are depicted in blue and positive

signals in red. The arrow labelled 'TC' highlights the transient crosspeak. **c**, Horizontal (red) and diagonal (blue) cuts through the transient 2D-IR spectra at the position shown by the red and blue dotted lines in **b**. The noise in absorbance change is $\Delta A = 1.5 \times 10^{-6}$ indicated by the error bar (± 1 s.d.) in the right panel.

'1' and '2' respectively in Fig. 2b, are stronger in the equilibrium 2D-IR spectrum), implying that it is a change of interaction strength, and not the overall frequency shift, that is responsible for the transient crosspeak.

To guide the interpretation of our experimental data, we simulate the photo-triggered conformational changes in our system using

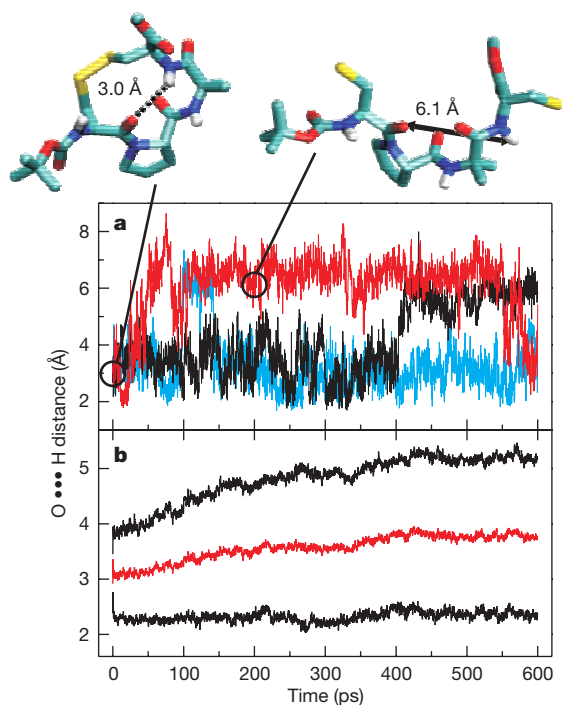


Figure 4 | Molecular dynamics results. **a**, O...H distance of three typical trajectories and two example structures. **b**, Mean O...H distance averaged over 120 non-equilibrium trajectories (red curve) together with ± 1 s.d. of the distribution of distances (black curves).

non-equilibrium molecular dynamics²² (see Supplementary Information for computational methods). Figure 4a shows the time evolution of the O...H distance after cleavage of the S-S bond for three typical non-equilibrium trajectories, and gives two simulated structures with short and long O...H distances. Figure 4b gives the mean O...H distance averaged over 120 non-equilibrium trajectories (red curve), and indicates the distribution of simulated distances at the one standard deviation (1 s.d.) level. We find that the mean O...H distance increases on a 240 ps timescale, in reasonable agreement with the experimentally observed timescale of 160 ps. However, the hydrogen bond does not break entirely. Instead, the distribution of O...H distances widens, with the fraction of trajectories characterized by short distances (corresponding to hydrogen-bonded molecules; see lower black curve in Fig. 4b) staying almost constant over time. This behaviour is consistent with NMR observations that during the synthesis of the peptide, the open, S-protected precursor molecule forms a weak hydrogen bond¹⁵. It is also consistent with the sudden jumps seen in the single trajectories (Fig. 4a), which are indicative of a two-state equilibrium between an open peptide conformation and a closed, hydrogen-bonded peptide conformation that are separated by a small energy barrier. Cleavage of the stabilizing S-S bond shifts the corresponding equilibrium constant from $K \approx 0.02$ to $K \approx 0.3$.

The molecular dynamics simulations indicate that, locally, the backbone structure between Cys(1)-Pro and between Cys(1)-Boc (crosspeaks '1' and '2' in Fig. 2b) does not change during the conformational transition (see the example structures in Fig. 4a), explaining why no net signal remains in a two-dimensional difference spectrum in the corresponding crosspeak region. In contrast, the relative orientation and distance of the C=O groups of Cys(1)-Aib (crosspeak '3' in Fig. 2b) change significantly during the conformational transition, giving rise to the transient crosspeak labelled 'TC' in Fig. 3. In other words, photocleavage of the disulphide bridge causes the rigid β -turn to adopt a more floppy and random structure that weakens the intramolecular hydrogen bond; this in turn manifests itself in a modified C=O coupling across the hydrogen bond.

We note that the β -turn opening we observe occurs on an ultrafast timescale with $k_{\text{obs}}^{-1} = 160$ ps. This rate is similar to what has been predicted from molecular dynamics simulations for the backbone dynamics of small peptides of similar length^{23,24}, and two orders of magnitudes faster than the speed limit established by triplet–triplet transfer methods¹⁴ for the contact formation between side chains in related molecules. As the equilibrium constant between open and closed peptide configurations is of the order of one after photocleavage of the S–S bond, the intrinsic rates for loop breaking k_{break} and loop formation k_{close} will be approximately the same with $k_{\text{obs}} = k_{\text{break}} + k_{\text{close}}$. Rates on this timescale can be considered an absolute speed limit for protein dynamics, and are governed solely by the relatively small intrinsic barriers of the peptide backbone and unaffected by entropic factors that become relevant in larger peptide chains or proteins^{14,25–27}. In this context, we note that peptide dynamics is expected to scale with solvent viscosity^{14,28}, which differs for the solvent CD₃CN we have used and the biologically relevant solvent water. Molecular dynamics simulations (data not shown) indeed suggest a change of the order of one in the timescale of the processes probed, as expected for a viscosity effect. However, a more serious problem can also arise from the fact that differences in solvent polarity and hydrogen-bonding characteristics can change which specific structures are most stable or most readily formed and also change the reaction or transformation pathways of the system. Nevertheless, transient 2D-IR experiments can be conducted using water as solvent^{2,3,9,10} (our choice of CD₃CN was simply due to the low water solubility of the peptide studied), and can in principle also readily be scaled to more complex systems. To resolve spectral congestion in larger molecules (for example, of the size of an α -helix) isotope labelling will be needed^{2,3}.

Transient 2D-IR spectroscopy combines picosecond time resolution with the ability to sense local contacts between specific molecular groups in the solution phase. Molecular dynamics simulations access similar length and timescales and can be used to investigate ultrafast non-equilibrium dynamical processes, making comparisons between results obtained with these two methods straightforward and useful for their mutual validation and interpretation. We thus anticipate that the combined use of transient 2D-IR spectroscopy and molecular dynamics simulations will deliver unprecedented insight into fast biomolecular processes.

Received 13 June; accepted 12 October 2006.

- Schotte, F. *et al.* Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. *Science* **300**, 1944–1947 (2003).
- Fang, C. & Hochstrasser, R. M. Two-dimensional infrared spectra of the ¹³C:¹⁸O isotopomers of alanine residues in an α -helix. *J. Phys. Chem. B* **109**, 18652–18663 (2005).
- Wang, J., Chen, J. & Hochstrasser, R. M. Local structure of β -hairpin isotopomers by FTIR, 2D IR, and ab initio theory. *J. Phys. Chem. B* **110**, 7545–7555 (2006).
- Asbury, J. B. *et al.* Hydrogen bond dynamics probed with ultrafast infrared heterodyne-detected multidimensional vibrational stimulated echoes. *Phys. Rev. Lett.* **91**, 237402 (2003).
- Mukamel, S. Multidimensional femtosecond correlation spectroscopies of electronic and vibrational excitations. *Annu. Rev. Phys. Chem.* **51**, 691–729 (2000).
- Mukherjee, P., Kass, I., Arkin, I. & Zanni, M. T. Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc. Natl Acad. Sci. USA* **103**, 3528–3533 (2006).
- Golonzka, O., Khalil, M., Demirdoven, N. & Tokmakoff, A. Vibrational anharmonicities revealed by coherent two-dimensional infrared spectroscopy. *Phys. Rev. Lett.* **86**, 2154–2157 (2001).

- Cowan, M. L. *et al.* Ultrafast memory loss and energy redistribution in the hydrogen bond network of liquid H₂O. *Nature* **434**, 199–202 (2005).
- Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102**, 6123–6138 (1998).
- Woutersen, S. & Hamm, P. Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy. *J. Phys. Chem. B* **104**, 11316–11320 (2000).
- Woutersen, S. & Hamm, P. Nonlinear two-dimensional vibrational spectroscopy of peptides. *J. Phys. Condens. Matter* **14**, R1035–R1062 (2002).
- Bredenbeck, J. *et al.* Transient 2D-IR spectroscopy: snapshots of the nonequilibrium ensemble during the picosecond conformational transition of a small peptide. *J. Phys. Chem. B* **107**, 8654–8660 (2003).
- Bredenbeck, J., Helbing, J. & Hamm, P. Transient two-dimensional infrared spectroscopy: exploring the polarization dependence. *J. Chem. Phys.* **121**, 5943–5957 (2004).
- Bieri, O. *et al.* The speed limit for protein folding measured by triplet–triplet energy transfer. *Proc. Natl Acad. Sci. USA* **96**, 9597–9601 (1999).
- Kolano, C., Gomann, K. & Sander, W. Small cyclic disulfide peptides: Synthesis in preparative amounts and characterization by means of NMR and FT-IR spectroscopy. *Eur. J. Org. Chem.* **20**, 4167–4176 (2004).
- Volk, M. *et al.* Peptide conformational dynamics and vibrational stark effects following photoinitiated disulfide cleavage. *J. Phys. Chem. B* **101**, 8607–8616 (1997).
- Lu, H. S. M. *et al.* Amino-thiopyran disulfide, an optical trigger for initiation of protein folding. *J. Am. Chem. Soc.* **119**, 7173–7180 (1997).
- Bredenbeck, J. *et al.* Picosecond conformational transition and equilibration of a cyclic peptide. *Proc. Natl Acad. Sci. USA* **100**, 6452–6457 (2003).
- Bredenbeck, J., Helbing, J., Kumita, J. R., Woolley, G. A. & Hamm, P. α -Helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved IR spectroscopy. *Proc. Natl Acad. Sci. USA* **102**, 2379–2384 (2005).
- Hamm, P., Ohline, S. M. & Zinth, W. Vibrational cooling after ultrafast photoisomerization of azobenzene measured by femtosecond infrared spectroscopy. *J. Chem. Phys.* **106**, 519–529 (1997).
- Pimentel, G. C. & McClellan, A. L. *The Hydrogen Bond* 67–141 (W. H. Freeman, San Francisco, 1960).
- Nguyen, P. H., Gorbunov, R. D. & Stock, G. Photoinduced conformational dynamics of a photoswitchable peptide: a nonequilibrium molecular dynamics simulation study. *Biophys. J.* **91**, 1224–1234 (2006).
- Mohanty, D., Elber, R., Thirumalai, D., Beglov, D. & Roux, B. Kinetics of peptide folding: computer simulations of SYPPDV and peptide variants in water. *J. Mol. Biol.* **272**, 423–442 (1997).
- Mohanty, D., Elber, R. & Thirumalai, D. Probing the role of local propensity in peptide turn formation. *Int. J. Quantum Chem.* **80**, 1125–1128 (2000).
- Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding ‘speed limit’. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).
- Sadqi, M., Fushman, D. & Munoz, V. Atom-by-atom analysis of global downhill protein folding. *Nature* **442**, 317–321 (2006).
- Yang, W. Y. & Gruebele, M. Folding at the speed limit. *Nature* **423**, 193–197 (2003).
- Satzger, H. *et al.* Picosecond dynamics in water-soluble azobenzene-peptides. *Chem. Phys. Lett.* **396**, 191–197 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was financially supported by a Postdoctoral Research Fellowship from the Deutsche Forschungsgemeinschaft to C.K.; by grants from the SNF to P.H.; and grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to W.S.

Author Contributions C.K. designed the molecule together with W.S., C.K. synthesized it and ran the experiment together with J.H. C.K., J.H. and P.H. wrote the manuscript. M.K. and P.H. were responsible for the molecular dynamics simulations. 2D-IR and transient 2D-IR spectroscopy is a major theme of the group of P.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.H. (p.hamm@pci.unizh.ch) or C.K. (c.kolano@pci.unizh.ch).

Slip zone and energetics of a large earthquake from the Taiwan Chelungpu-fault Drilling Project

Kuo-Fong, Ma¹, Hidemi Tanaka², Sheng-Rong Song³, Chien-Ying Wang¹, Jih-Hao Hung¹, Yi-Ben Tsai¹, Jim Mori⁴, Yen-Fang Song⁵, Eh-Chao Yeh⁶, Wonn Soh⁶, Hiroki Sone⁷, Li-Wei Kuo³ & Hung-Yu Wu¹

Determining the seismic fracture energy during an earthquake and understanding the associated creation and development of a fault zone requires a combination of both seismological and geological field data¹. The actual thickness of the zone that slips during the rupture of a large earthquake is not known and is a key seismological parameter in understanding energy dissipation, rupture processes and seismic efficiency. The 1999 magnitude-7.7 earthquake in Chi-Chi, Taiwan, produced large slip (8 to 10 metres) at or near the surface², which is accessible to borehole drilling and provides a rare opportunity to sample a fault that had large slip in a recent earthquake. Here we present the retrieved cores from the Taiwan Chelungpu-fault Drilling Project and identify the main slip zone associated with the Chi-Chi earthquake. The surface fracture energy estimated from grain sizes in the gouge zone of the fault sample was directly compared to the seismic fracture energy determined from near-field seismic data^{3,4}. From the comparison, the contribution of gouge surface energy to the earthquake breakdown work is quantified to be 6 per cent.

The North–South-trending Chelungpu fault is a major 90-km structure that dips shallowly to the east (30°), and principally slips within, and parallel to, bedding of the Pliocene Chinshui shale⁵. Taiwan Chelungpu-fault Drilling Project (TCDP) drilled two vertical holes 40 m apart (hole A to a depth of 2 km, and hole B to a depth of 1.3 km), and a side-track from hole B (hole C) at the depth of 950 m to 1,200 m about 2 km east of the surface rupture, near the town of DaKeng (Fig. 1a). The subsurface location of the Chinshui shale was known from high-resolution seismic reflection profiles^{6,7} at a depth of about 1,000 m under the DaKeng site. The spatial slip distribution for the earthquake was well constrained from close strong motion stations and Global Positioning System (GPS) data^{3,4} and shows a slip of 8.3 m on the fault near the drill site. The drilling carried out continuous coring for depths of 500–2,000 m for hole A, 950–1,300 m for hole B and 950–1,200 m for hole C, respectively. Geophysical well logs were carried out in hole A to collect seismic velocities, densities and digital images.

From the hole-A core, the Chelungpu fault zone is seen within the Chinshui shale as a damaged zone at depths of about 1,105 to 1,115 m, consisting of fault breccia and fault gouge (Fig. 1b). The degree of fracturing increases from the top to the bottom of the zone. Near the bottom of the broad zone of deformation, a 12-cm-thick primary slip zone (PSZ) can be identified based on the presence of ultra-fine-grained fault gouge and increased fracture density at depths of 1,111.23 to 1,111.35 m. A corresponding feature was also found in the hole-B core at depths of 1,136.50 to 1,136.62 m, confirming the fault dip of 30° E. The geophysical logging measurements of low

seismic velocities and low electrical resistivity around the depth of 1,111 m also confirm that this is the main fault zone.

The PSZ seen in the core from hole C after splitting and polishing (Fig. 1c), shows several layers of slip zones associated with several repeating earthquakes. The individual slip zone has a thickness of about 2–3 cm with a 5-mm ultrafine grain zone in the bottom as indicated in the PSZ schematic (Fig. 1c). Among the slip zones, the least deformed region, which has the fewest number of cross-cutting cracks, is the 2-cm zone at the bottom of the PSZ, suggesting that this narrow band might be the major slip zone (MSZ) that corresponds to the Chi-Chi earthquake. Other estimates of the thickness for the slip zone from nearby sites are 50–300 μm observed at the surface near the DaKeng drill site⁸, and 7 mm from a fault core at a depth of 330 m in shallow drilling before TCDP⁹. These determinations of slip zone thicknesses are all from layers located near the bottom of the fracture zone. The variation of thickness of the slip zone at different depths might correspond to differences in normal stress^{10–12}.

We also analysed the grain size distribution of the slip zone^{13,14} using transmission electron microscope (TEM), scanning electron microscope (SEM) and optical microscope measurements, to estimate the surface fracture energy associated with the gouge formation. The distribution of particle size is shown in Fig. 2a, which follows a power-law distribution with a slope of about 2.3 (refs 15 and 16; see Supplementary Information). Grain sizes of 50 nm–100 μm (Fig. 2b, Supplementary Fig. 1a–c) were observed for the 2-cm MSZ (Supplementary Table 1). We consider grain sizes larger than 50 nm for the surface fracture energy calculation. The images with grain sizes of less than 50 nm show rounded shapes, suggesting that those small grains might be the result of precipitation rather than fracturing (Fig. 2b). Assuming spherical grains and a ratio of surface area to volume for spheres of 3/radius (ref. 13), we obtained the total particle surface area for the 2-cm slip zone S_{MSZ} of $6.46 \times 10^5 \text{ m}^2$ per metre squared area. The mineral composition from X-ray diffraction for semiquantitative analysis shows that the MSZ was composed of about 70% of quartz, 5% of feldspar, and 25% of clay minerals (Supplementary Fig. 2). This gives a specific fracture energy G_c of about 1 J m^{-2} (refs 17–19). Using a correction for grain roughness λ of 6.6 (ref. 20), and the specific fracture energy, we obtain the surface fracture energy G_{MSZ} of the 2-cm MSZ by:

$$G_{\text{MSZ}} = S_{\text{MSZ}} \lambda G_c \quad (1)$$

From equation (1), we obtain a value of 4.3 MJ m^{-2} for the surface fracture energy. This is interpreted as the minimum amount of energy that is necessary to produce the MSZ in one earthquake.

¹Department of Earth Sciences, National Central University, Chung-Li 32054, Taiwan. ²Department of Earth and Planetary Sciences, University of Tokyo, Tokyo 113-0033, Japan. ³Department of Geosciences, National Taiwan University, Taipei 10617, Taiwan. ⁴Disaster Prevention Research Institute, Kyoto University, Kyoto 611-0011, Japan. ⁵National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan. ⁶Kochi Institute for Core Sample Research, Agency for Marine–Earth Science and Technology, Kochi 783-8502, Japan. ⁷Department of Geology and Mineralogy, Kyoto University, Kyoto 606-8501, Japan.

Any additional contribution to the estimated surface fracture energy from considering a much wider, but less deformed, damage zone is thought to be less than 10% of the above value¹³.

Formation of the slip zone is associated with the seismic fracture energy, which is consumed as the earthquake rupture propagates. The creation of small grains is a contribution, but may not be the total equivalent to the seismic fracture energy. To estimate the seismic fracture energy from the observed earthquake waves, we used a three-dimensional finite difference code based on the traction-at-split nodes method to calculate the stress time history on the earthquake fault plane²¹, with the temporal-spatial slip distribution inverted from strong motion data⁴ as a constraint. The seismic fracture energy density on the fault is determined by retrieving the dynamic traction evolution during the slip history.

The breakdown work²² W_b , which is the excess work over some minimum level achieved during slip, is the energy spent to allow the rupture to advance. It can be obtained from the slip history of shear traction on the fault, by calculating the integral of the traction versus slip, from zero slip to the point that the traction drops to a minimum:

$$W_b = \int_0^{T_b} (\tau(t) - \tau_{\min}) \cdot \mathbf{v}(t) dt \quad (2)$$

where $\mathbf{v}(t)$ is the slip velocity, $\tau(t)$ is the shear traction, and T_b is the time at which minimum traction $\tau_{\min}$ is reached. A grid size of 0.95 km and a time interval of 0.054 s were used for the calculation. Figure 3 shows the shear traction change as a function of slip and time for the particular portion of the fault beneath the borehole site. The shaded area in Fig. 3 corresponds to the integral in equation (2), and gives a value for the breakdown work W_b of 11.6 MJ m^{-2} . This value is comparable to that for other nearby subfaults, and the results of other studies²³. This breakdown work can be considered as an

equivalent to the seismic fracture energy density²² G_s . The breakdown work derived from the equation (2) is composed of the surface fracture energy for formation of the slip zone, and other dissipative losses during faulting²². Here, we assume that the fault core thickness, fault geometry, clast grain size and distribution retrieved from TCDP do not significantly change over the subfault area of the seismic inversion.

The geological studies^{5,8} show that the total displacement accommodated by the Chelungpu–Chinshui detachment, where the TCDP drilled through, is 0.3 km. Considering the 12-cm primary slip zone identified from the retrieved core, the ratio (T/D) of the slip thickness ($T = 12 \text{ cm}$) to the total displacement ($D = 300 \text{ m}$) is 4×10^{-4} . For the 8.3-m slip of the Chi-Chi earthquake, the slip thickness for a single earthquake is 3.3 mm. This means that the number of events in the 2-cm MSZ is between 6 and 7 if we assume similar displacements of repeating earthquakes in the major slip zone. For the 4.3 MJ m^{-2} of the surface fracture energy from the 2-cm MSZ, the fracture surface energy associated to a single earthquake on average would be about 0.65 MJ m^{-2} . Given that the breakdown work is 11.6 MJ m^{-2} , this value shows that the process of grain formation represents about 6% of the earthquake breakdown work. We consider this estimate to be the maximum for the assumption that there is no fracture energy occurred during sliding¹³. The remaining part of the breakdown work will mostly be heat, which might be associated with other dynamic processes, such as fault thermopressurization^{24–26}, or fault lubrication^{27,28}.

The radiation efficiency η_R is the ratio of the radiated energy E_R to the energy available for mechanical processes. It was defined²⁹ as $\eta_R = \frac{E_R}{E_R + E_G}$, where E_G is the product of fault area and seismic fracture energy density. Here, we consider the seismic fracture energy in the equation to be the surface energy used to pulverize the rock for

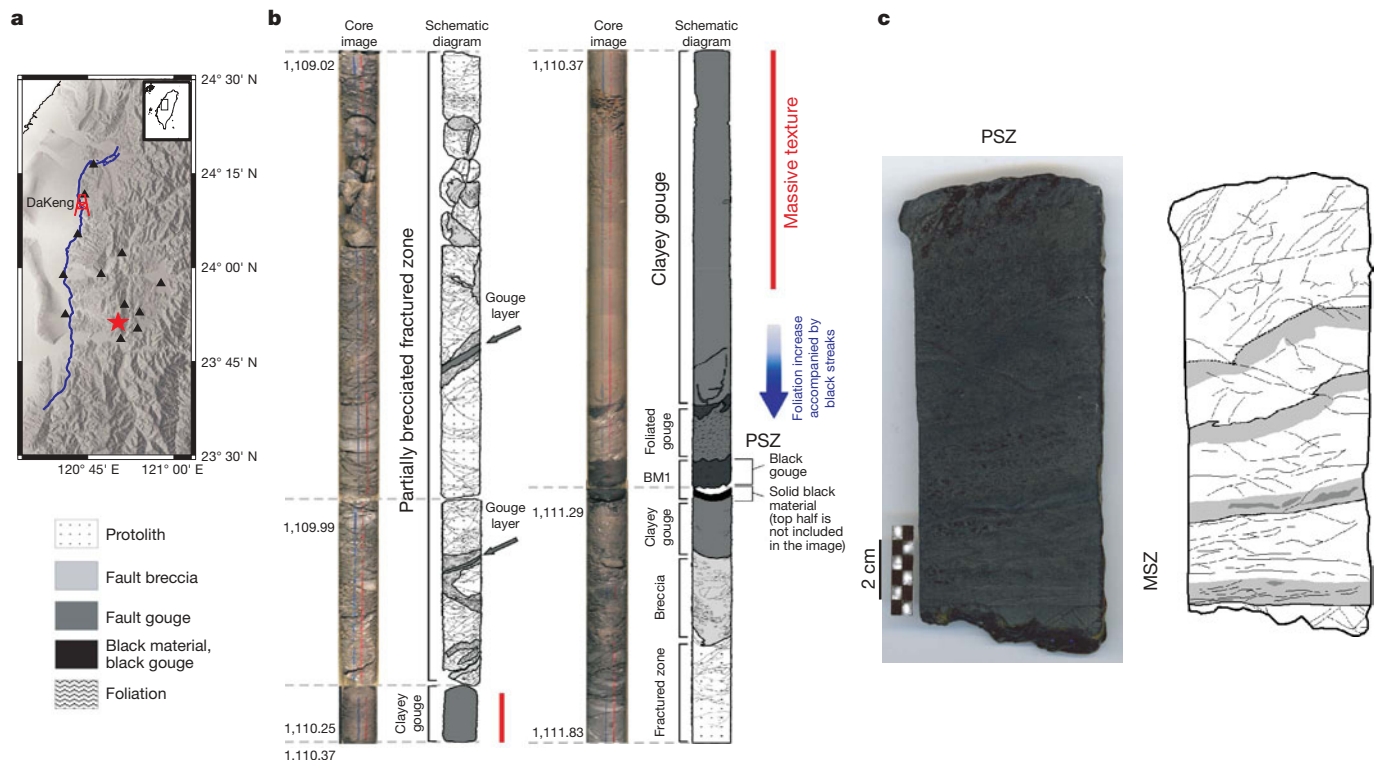


Figure 1 | Location, core images and polished primary slip zone. **a**, The location of the drill site at the town DaKeng and the ruptured Chelungpu fault (bold blue line). The black triangles show the distribution of the strong motion stations close to the fault. The epicentre is shown by a red asterisk. **b**, The core image and schematic diagram from the depths of 1,109.02 m to 1,111.83 m of hole A with descriptive comments. A 12-cm primary slip zone was observed at the depth of 1,111.23–1,111.35 m. **c**, An enlarged photo of

the splitting and polishing slab of the 12-cm principal slip zone (PSZ) with its schematic. The thicker lines in the schematic indicate the possible slip zones associated with several repeating earthquakes. The 5-mm ultrafine grain zone in the bottom of each layer is shown in grey. The bottom layer with the less-deformed slip zone is the major slip zone (MSZ) related to the 1999 Chi-Chi earthquake.

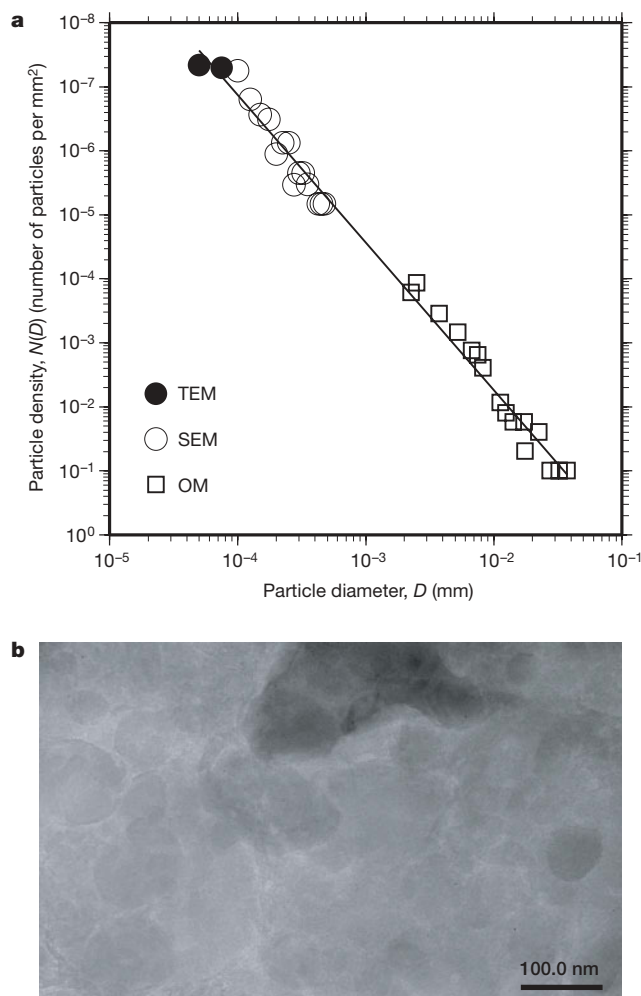


Figure 2 | Particle size in the major slip zone. **a**, Distribution of the particle size, $N(D)$ as a function of particle diameter (D) in millimetres. The $N(D)$ is the number of grains per mm^2 for a class of grain size. The measurements are imaged from TEM (solid circles), SEM (circles), and optical (square). The regression of the particle size distribution follows the power law $N(D) = aD^{-b}$, where a is 0.0045 and b is 2.3. **b**, TEM image; scale, 100 nm.

formation of the fault gouge. Assuming that the ratio of the seismic energy to seismic moment is constant, η_R can be modified to the quantity¹:

$$\eta'_R = \frac{1}{1 + \frac{6\lambda}{\mu C_R} \left(\frac{T}{D}\right) \left(\frac{G_c}{\hat{d}}\right)} \quad (3)$$

where $C_R = E_R/M_0$, and M_0 is the seismic moment. The value of C_R is 2×10^{-5} , using values of E_R and M_0 for the Chi-Chi earthquake derived from seismic data^{4,30}. The representative grain size $\hat{d}$ has a value of 186 nm from $\hat{d} = 6T/S_{\text{MSZ}}$, which has an amount of surface fracture energy equivalent to that of the power-law size-distribution in the MSZ. If the surface energy is the major contribution to the fracture energy, η'_R will have a similar value to η_R . For a rigidity $\mu = 30$ GPa, we obtain a value of η'_R of 0.88 for the Chlungpu fault (Fig. 4).

The gouge zone of the Chlungpu fault for the Chi-Chi earthquake was formed as a result of about 6% of the breakdown work. The calculated value for η'_R is intermediate between that of the well-developed Punchbowl fault in California¹³ ($\eta'_R \approx 1.0$) and those of the earthquakes in a South African mine that are ($\eta'_R \approx 0.16$) associated with making new fractures¹⁸. The physical differences in the fault zones may depend on the maturity and style of faulting¹³ and reflect the differences in the mechanical energy absorbed during large

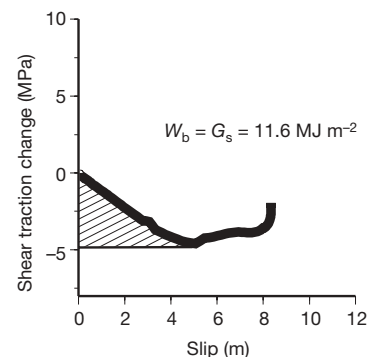


Figure 3 | The slip-weakening curve for the fault block corresponds to the borehole site. The shaded area gives an estimate for W_b , equivalent to G_s , of about 11.6 MJ m^{-2} .

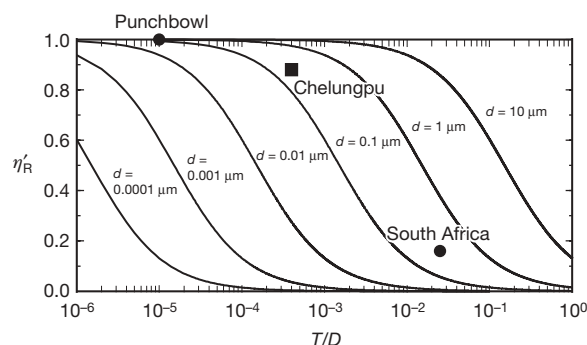


Figure 4 | The ratio of radiated energy to the summation of the radiated and surface energy as a function of the ratio of fault thickness (T) to total fault displacement (D). Curves show the values for various grain sizes. The values for the Punchbowl fault, South African mines (solid circles) and the Chlungpu fault (squares) are shown. The Chlungpu fault has intermediate values between the Punchbowl fault (1.0) and data from the South African mines (0.16).

events. When large earthquakes occur on mature faults, there is less fracturing, so the proportional amount of dissipative energy is smaller, compared to the more brittle behaviour of young faults.

Received 28 February; accepted 13 September 2006.

- Kanamori, H. The diversity of the physics of earthquakes. *Proc. Jpn. Acad.* **80**, 297–316 (2004).
- Ma, K.-F. *et al.* The Chi-Chi, Taiwan earthquake: large surface displacements on an inland thrust fault. *Eos* **80**, 605 (1999).
- Ma, K.-F. *et al.* Spatial and temporal distribution of slip for the 1999 Chi-Chi, Taiwan earthquake. *Bull. Seismol. Soc. Am.* **91**, 1069–1087 (2001).
- Ji, C. *et al.* Slip history dynamic implications of the 1999 Chi-Chi, Taiwan, earthquake. *J. Geophys. Res.* **108**, 2412, doi:10.1029/2002JB001764 (2003).
- Yue, L. F., Suppe, J. & Hung, J.-H. Structure geology of a classic thrust belt earthquake: the 1999 Chi-Chi earthquake Taiwan (Mw=7.6). *J. Struct. Geol.* **27**, 2058–2083 (2005).
- Wang, C.-Y. Detection of a recent earthquake fault by the shallow reflection seismic method. *Geophysics* **67**, 1465–1473 (2002).
- Wang, C.-Y., Li, C.-L. & Yen, H.-Y. Mapping the northern portion of the Chlungpu fault, Taiwan by shallow reflection seismics. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014496 (2002).
- Heermance, R., Shipton, Z. K. & Evans, J. P. Fault structure control on fault slip and ground motion during the 1999 rupture of the Chlungpu fault, Taiwan. *Seismol. Soc. Am. Bull.* **93**, 1034–1050 (2003).
- Tanaka, H. *et al.* Initial science report of shallow drilling penetrating into the Chlungpu fault zone, Taiwan. *Terr. Atmos. Ocean. Sci.* **113**, 227–251 (2002).
- Sibson, R. H. Thickness of the seismic slip zone. *Bull. Seismol. Soc. Am.* **93**, 1169–1178 (2003).
- Tanaka, H. *et al.* Whole fault zone architecture and its relation to thin slip layer activated by 1995 Kobe earthquake detected at 1140 m depth in the drilled core penetrating the Nojima fault. *J. Geophys. Res.* (submitted).
- Gratier, J.-P., Favreau, P. & Renard, F. Modeling fluid transfer along California faults when integrating pressure solution crack sealing and compaction process. *J. Geophys. Res.* **108**, doi:10.1029/2001JB000380 (2003).

13. Chester, J. S., Chester, F. M. & Kronenberg, A. K. Fracture surface energy of the Punchbowl Fault, San Andreas system. *Nature* **437**, 133–136 (2005).
14. Sammis, C., King, G. & Biegel, R. The kinematics of gouge deformation. *Pure Appl. Geophys.* **125**, 777–812 (1987).
15. Otsuki, K., Monzawa, N. & Nagase, T. Fluidization and melting of fault gouge during seismic slip: Identification in the Nojima fault zone and implications for focal earthquake mechanisms. *J. Geophys. Res.* **108**, B4 2192, doi:10.1029/2001JB001711 (2003).
16. Kadono, T. *et al.* Surface roughness of alumina fragments caused by hypervelocity impact. *Planet. Space Sci.* **54**, 212–215 (2006).
17. Scholz, C. H. in *The Mechanics of Earthquakes and Faulting* 158–167 (Cambridge Univ. Press, Cambridge, UK, 2002).
18. Lawn, B. *Fracture of Brittle Solids* 2nd edn 1–378 (Cambridge Univ. Press, New York, 1993).
19. McGarr, A., Spottiswoode, S. M. & Gay, N. C. Observations relevant to seismic driving stress, stress drop, and efficiency. *J. Geophys. Res.* **84**, 2251–2261 (1979).
20. Wilson, B., Dewers, T., Reches, Z. & Brune, J. Particle size and energetics of gouge from earthquake rupture zones. *Nature* **434**, 749–752 (2005).
21. Andrews, D. J. Test of two methods for faulting in finite-difference calculations. *Bull. Seismol. Soc. Am.* **89**, 4931–4937 (1999).
22. Tinti, E., Spudich, P. & Cocco, M. Earthquake fracture energy inferred from kinematic rupture models on extended faults. *J. Geophys. Res.* **110**, B12303 doi:10.1029/2005JB003644 (2005).
23. Zhang, W. B. *et al.* Heterogeneous distribution of the dynamic source parameters of the 1999 Chi-Chi, Taiwan, earthquake. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001889 (2003).
24. Abercrombie, R. E. & Rice, J. R. Can observation of earthquake scaling constrain slip weakening? *Geophys. J. Int.* **162**, 406–424 (2005).
25. Wibberley, C. A. J. & Shimamoto, T. Earthquake slip weakening and asperities explained by thermal pressurization. *Nature* **436**, 689–692 (2005).
26. Tanaka, H. *et al.* Frictional heat from faulting of the 1999 Chi-Chi, Taiwan earthquake. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL026673 (2006).
27. Brodsky, E. & Kanamori, H. Elastohydrodynamic lubrication of faults. *J. Geophys. Res.* **106**, 16357–16373 (2001).
28. Ma, K.-F. *et al.* Evidence for fault lubrication during the 1999 Chi-Chi, Taiwan, earthquake (Mw7.6). *Geophys. Res. Lett.* **30**, 1244 doi:10.1029/2002GL015380 (2003).
29. Hussein, M. I. & Randall, M. J. Rupture velocity and radiation efficiency. *Seismol. Soc. Am. Bull.* **660**, 1173–1187 (1976).
30. Venkataraman, A. & Kanamori, H. Effect of directivity on estimates of radiated seismic energy. *J. Geophys. Res.* **109**, doi:10.1029/2003JB002548 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the working group of TCDP, including the drilling company FangYu and WonDa, the on-site assistants and more than 60 participating students from NCU and NTU, S. T. Huang at CPC for core storage and splitting, Y.-M. Chen of the NSRRC for taking TEM and TXM images, and K. S. Liang of the NSRRC. We also thank M. Zoback, S. Hickman, W. Ellsworth and H. Ito for discussion before and during drilling. We also thank H. Kanamori and E. Brodsky for discussions and a review of an early version of this paper. This project was supported by the NSC, Taiwan, and partially supported by the ICDP.

Author Contributions K.-F.M., paper writing and project planning. H.T., core observation, data analysis and project planning. S.-R.S., petrographic and XRD analysis and project planning. C.-Y.W., J.-H.H., Y.-B.T., W.S., project planning. J.M., paper preparation. Y.-F.S., TEM and TXM taken at NSRRC and data analysis. E.-C.Y. and H.S., on-site geologists during coring. L.-W.K. and H.-Y.W., on-site assistants.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.-F.M. (fong@earth.ncu.edu.tw).

Ecological consequences of major hydrodynamic disturbances on coral reefs

Joshua S. Madin^{1,2} & Sean R. Connolly¹

A recent tsunami^{1,2} and an apparent increase in the frequency of severe tropical storms^{3–5} underscore the need to understand and predict the ecological consequences of major hydrodynamic disturbances⁶. Reef corals provide the habitat structure that sustains the high biodiversity of tropical reefs⁷, and thus provide the foundation for the ecosystem goods and services that are critical to many tropical societies^{6,8}. Here we integrate predictions from oceanographic models⁹ with engineering theory, to predict the dislodgement of benthic reef corals during hydrodynamic disturbances. This generalizes earlier work^{10,11}, by incorporating colonies of any shape and by explicitly examining the effects of hydrodynamic gradients on coral assemblage structure. A field test shows that this model accurately predicts changes in the mechanical vulnerability of coral colonies, and thus their size and shape, with distance from the reef crest. This work provides a general framework for understanding and predicting the effects of hydrodynamic disturbances on coral reef communities; such disturbances have a major role in determining species zonation^{12,13} and coexistence¹⁴ on coral reefs, and are critical determinants of how coral assemblages will respond to changes in the frequency and intensity of tropical storms associated with a changing climate^{3–5,15}.

A framework to predict whole-colony dislodgement of corals caused by hydrodynamic force requires three elements^{16,17}: (1) engineering theory for translating levels of ambient water motion into internal mechanical stress within the coral; (2) measures of the maximum stress that colonies can withstand before breakage (that is, material strength); and (3) estimates of the probability that the maximum tolerable stress is exceeded where colonies live on the reef. For corals, internal tensile stress at the base and periphery of a colony—produced by hydrodynamic drag force (see Supplementary Information)—limits mechanical integrity in the vast majority of cases¹⁸. Because reef substrate is substantially weaker than coral skeleton (particularly for morphologies prone to hydrodynamic dislodgement), the capacity to resist dislodgement by a hydrodynamic force is generally limited by the tensile strength of the reef substrate¹⁸. Therefore, if the basal tensile stress produced by a hydrodynamic event exceeds the substrate strength, dislodgement is expected. Using cantilever beam theory, and approximating the basal cross-sectional area of a colony as an ellipse, we can derive a dimensionless expression for the threshold mechanical vulnerability of colonies:

$$\frac{\sigma_s}{U^2 \rho_w} \leq \frac{16}{d_{||}^2 d_{\perp} \pi} \int_0^h y w(y) dy \quad (1)$$

(see Methods for derivation). U is horizontal water velocity, σ_s is the tensile strength of the substrate, ρ_w is water density, $d_{||}$ is the width of the base of the colony parallel to water flow, $d_{\perp}$ is the width of the

base perpendicular to water flow, y is distance above the substrate, and h is the height of the colony. $w(y)$ is the projected width (that is, excluding interstitial space between branches) of the colony perpendicular to water flow as a function of distance y above the substrate. The left side of the inequality is a dimensionless quantity describing the mechanical threshold imposed by the environment, which we term the Dislodgement Mechanical Threshold (*DMT*). The right side is a dimensionless quantity describing the colony's mechanical vulnerability, which we term the Colony Shape Factor (*CSF*). Dislodgement occurs when a colony's *CSF* exceeds the *DMT* imposed by a hydrodynamic event.

To test the model, we measured *CSF* for 1158 colonies from three coral species on an exposed coral reef at Lizard Island in the northern region of the Great Barrier Reef (see Methods). These species (*Acropora hyacinthus*, *Acropora gemmifera* and *Acropora palifera*) occur broadly along hydrodynamic gradients, and they have very different colony shapes. Thus, they are likely to exhibit substantial differences in *CSF*, and consequently to be ideal for testing a model of structural stability in reef corals. Using an existing 37-yr hindcast of hourly horizontal water velocity generated by waves over the study reef at 4-m increments⁹, and an existing *in situ* measurement of the average tensile strength of reef substrate for the study site¹⁸ ($\sigma_s = 0.2 \text{ MN m}^{-2}$) we used expression (1) to compare the mechanical threshold (*DMT*) imposed by the maximum hydrodynamic event during the year before the field test with the *CSFs* of colonies on the reef. This predicted threshold is highly consistent with field data on the distribution of *CSF* in our study species: *CSFs* are at or below the threshold for all species (Fig. 1). Indeed, *A. hyacinthus* provides exceptionally strong support for the model: the predicted *DMT* threshold corresponds extremely well with the maximum observed *CSF* values, closely following the gradient in maximum *CSF* with distance from the reef crest (solid line, Fig. 1a).

The differences among species in the extent to which they approach the predicted threshold (for example, *A. hyacinthus* abutting the threshold, but *A. palifera* remaining well below it) are also consistent with the model. This is apparent from differences among species in how colonies change shape with increasing size (Fig. 2). For *A. hyacinthus*, *CSF* increases with increasing colony size (Spearman's $\rho = 0.733$, $P < 0.001$, $n = 496$), so that colonies grow into increasing categories of vulnerability (Fig. 2a). For *A. gemmifera*, colonies beyond ~25 m of the reef crest approach the threshold more closely than colonies within 25 m of the crest (Fig. 1b). This pattern is caused by differences in how colonies of this species change shape as they grow. Towards the reef back, *A. gemmifera* adopt a tabular growth form, and become less stable as they grow (Fig. 2b, solid points; Spearman's $\rho = 0.491$, $P < 0.001$, $n = 311$). Consequently, like *A. hyacinthus*, the largest colonies approach the threshold. In contrast, near the crest, adults adopt a more digitate growth form, so they

¹School of Marine and Tropical Biology, and ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia. ²National Centre for Ecological Analysis and Synthesis, University of California, Santa Barbara, California 93101, USA.

retain low CSF values with increasing size (Fig. 2b, open points; Spearman's $\rho = 0.227$, $P = 0.003$, $n = 176$), and do not approach the threshold shown in Fig. 1. Finally, *A. palifera* exhibits pronounced increases in attachment area relative to projected area as a colony grows, which actually causes CSFs in this species to decrease as colony size increases (Spearman's $\rho = -0.258$, $P < 0.001$, $n = 175$; Fig. 2c). Consequently, like the digitate form of *A. gemmifera*, they never approach the threshold.

Using the largest hydrodynamic event captured in the 37-yr database (tropical cyclone Rona, a category 3 on the Saffir-Simpson scale, which passed 200 km to the South in 1999), we have also projected a plausible mechanical threshold produced by a 'moderate' cyclone (dashed line, Fig. 1). Our predictions indicate that many larger colonies from the *A. hyacinthus* and *A. gemmifera* populations would have been dislodged, but other large colonies would have remained, owing to spatial differences in maximal water velocity over the reef, as well as variation in how CSF changes with size for colonies that adopt different growth forms (Fig. 2, red points). The difference between the Rona and the 2002 thresholds also helps to explain why a large number of *A. hyacinthus* colonies approach the 2002 threshold so closely. Specifically, published estimates of colony growth for *A. hyacinthus* colonies, and increases in CSF with increasing colony size (Fig. 2a), indicate that *A. hyacinthus* colonies could readily double their CSFs within 1–2 yr, an increase comparable in magnitude to the difference between the two thresholds shown in Fig. 1a (see Supplementary Information for calculation). Thus, one would expect the largest *A. hyacinthus* colonies to grow beyond (and be removed

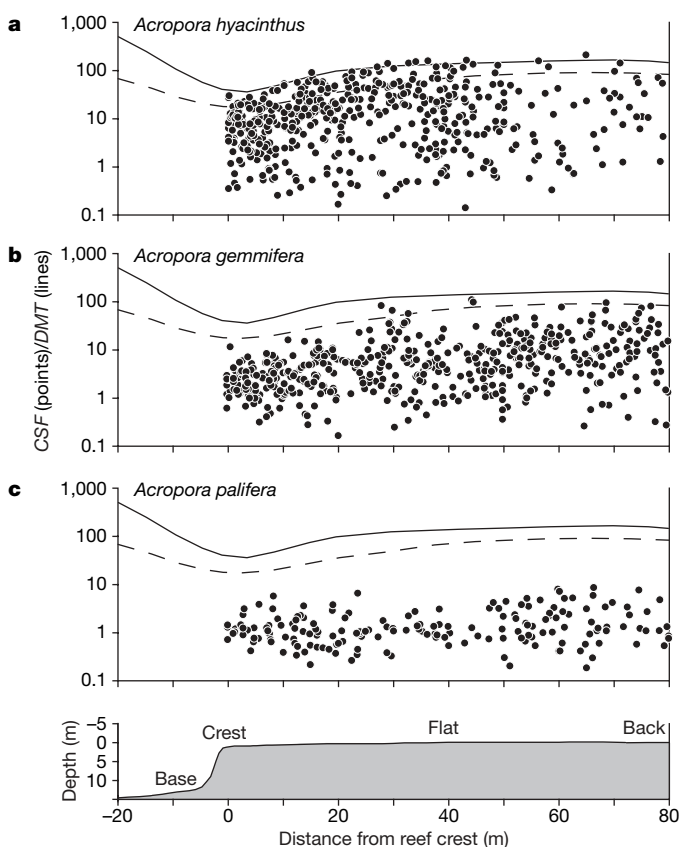


Figure 1 | Colony CSF as a function of colony distance from the reef crest. **a**, *A. hyacinthus*; **b**, *A. gemmifera*; **c**, *A. palifera*. Calculated CSFs are represented as points. The prediction of the DMT expected to have been imposed by the largest hydrodynamic event in the year before the field test (September 2002; see text for details) is represented by a solid line. The dashed line is the predicted DMT for Cyclone Rona (February 1999)—the most significant hydrodynamic event captured in the 37-yr oceanographic hindcast of the test site (see text for details). The reef depth profile is shown in the bottom panel.

by) annual thresholds in most years, and thus for these thresholds to correspond closely with the maximum observed CSF for this species, as in Fig. 1a. In contrast, the slower growing and less mechanically vulnerable *A. gemmifera* would probably take many times longer, on average, for comparable increases in CSF to occur (see Supplementary Information for calculation). Thus, even though tabular forms of *A. gemmifera* become more vulnerable as they grow, one would expect fewer colonies to reach the threshold for dislodgment in any given year, relative to *A. hyacinthus*, as Fig. 1 suggests.

Calibrating a model of structural stability for corals makes possible the estimation of colony dislodgement rates from time series of hydrodynamic disturbances. Specifically, maximum water velocities on reefs are wave-driven, and waves within the GBR lagoon are predominantly wind-generated and thus implicitly related to regional meteorological conditions⁹. Therefore, we should be able to approximate rare hydrodynamic disturbances and resulting colony dislodgement as a Poisson process¹⁹. To test this, we used the 37-yr database of water velocities and fitted exponential distributions to the frequency distributions of the waiting times in years between the rare wave events that would theoretically dislodge a colony of a given CSF. This analysis confirms that an exponential distribution closely approximates the distribution of waiting times (see Supplementary Information), and it allows us to calculate (from the fitted rate parameter of the exponential distribution, μ) estimates of annual

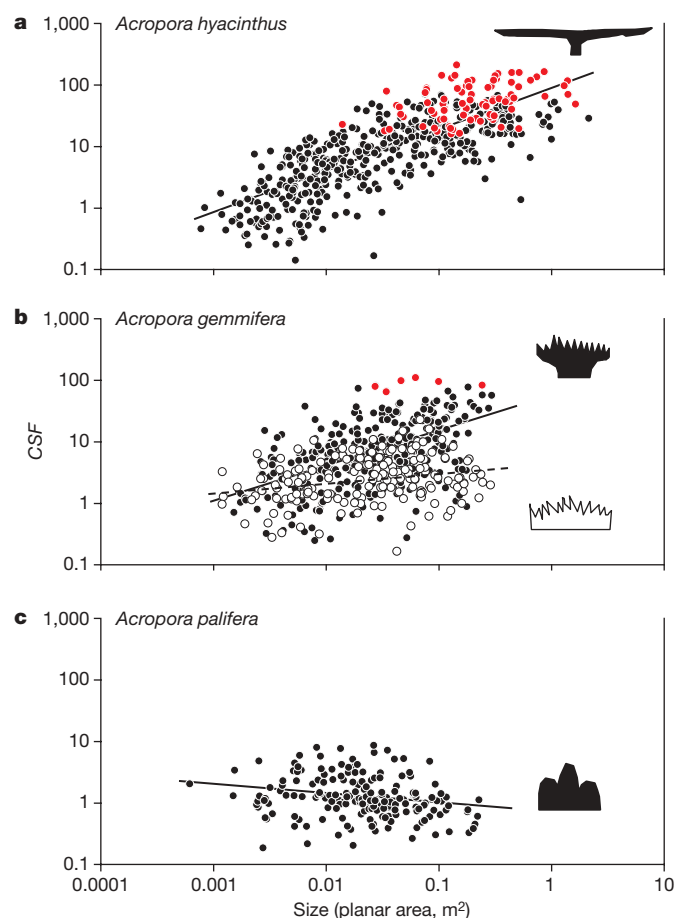


Figure 2 | Colony CSF as a function of colony size for three morphologically different coral reef species. **a**, *A. hyacinthus*; **b**, *A. gemmifera*; **c**, *A. palifera*. Least-squares regression lines are shown to illustrate trends with size. Silhouettes are examples of projected shapes for larger colonies. For *A. gemmifera*, open points indicate colonies living between the reef crest and 25 m, and solid points indicate colonies living beyond 25 m of the crest. Dashed and solid regression lines, respectively, indicate differences in how these colonies change shape with increasing size. Red points indicate the colonies above the estimated DMT in Fig. 1 for Cyclone Rona.

colony dislodgement caused by wave-induced disturbances on a reef platform, for colonies of any size and shape (Fig. 3). Specifically, if μ is the rate of occurrence of wave events whose threshold exceeds some colony's *CSF*, then the corresponding annual probability of dislodgement is $1 - e^{-\mu}$ (ref. 20). Colonies with high *CSF* (that is, mechanically vulnerable colonies) are dislodged by smaller waves that propagate further over the reef, and thus their mortality risk decreases gradually with distance from the crest. Indeed, for particularly vulnerable forms, such as those with a *CSF* of 250 or 500, annual mortality risk is approximately constant over much of the reef platform. In contrast, colonies with low *CSF* are dislodged by only very severe events. Because those events are associated with large waves, and because large waves break and attenuate rapidly^{9,21}, mortality risk for low *CSF* colonies drops off very rapidly with distance behind the reef crest.

Understanding how coral assemblages respond to hydrodynamic disturbances is essential for explaining zonation on present-day coral reefs, as well as for projecting how coral reef communities will change

in response to predicted increases in the frequency and intensity of tropical storms^{3–5}. A major obstacle to achieving both objectives has been our inability to estimate the effects of major disturbances on the mortality rates of different coral species. Whole-colony dislodgement occurs during episodic and very rare events. Consequently, direct estimates of mortality during such events can come from only the very few existing long-term monitoring studies that record abundances at the species level²² and our ability to extrapolate from these studies is constrained, because logistical considerations limit them to relatively small sample sizes and spatial scales. Our study offers a solution to this longstanding problem: by calculating the *CSF* of key reef coral species, the strength of the substrate to which they attach, and the attenuation of waves across reef platforms, we can predict the changes in size structure and species composition of coral assemblages caused by recent and future hydrodynamic disturbances. Such a framework is essential for understanding the dynamics of coral assemblages over space and time^{12–14}. This understanding can, in turn, facilitate informed policy and management decisions that aim to protect the world's coral reefs^{15,23}.

METHODS

Cantilever beam theory. For a given horizontal water velocity, the drag force F acting on an attached colony is directly proportional to the colony's area, A , projected onto a plane perpendicular to the direction of water flow¹⁶:

$$F = \left(\frac{\rho_w C_d}{2} \right) U^2 A \quad (2)$$

(see Supplementary Information). U is water velocity, ρ_w is water density (approximately 1025 kg m^{-3} for seawater) and C_d is the drag coefficient (approximately 1.0 for corals; see Supplementary Information). There is neither sufficient distance nor time as an orbital wave passes for a significant vertical velocity gradient to form at the scale of a colony (see Supplementary Information). Therefore, we assume that water motion acts similarly on all regions of a colony's projected area, in which case the maximum bending moment M is:

$$M = \frac{\rho_w U^2}{2} \int_{y=0}^h y w(y) dy \quad (3)$$

where h is colony height, y is vertical distance from the substrate, $w(y)$ is the projected width (that is, excluding interstitial space) of the colony perpendicular to water flow as a function of height y above the substrate.

The maximal tensile stress, σ_t , produced by bending a coral colony is given by²⁴:

$$\sigma_t = \frac{M d_{||}}{2 I_{xx}} \quad (4)$$

where M is the bending moment, $d_{||}$ is the width of the base of the colony parallel to water flow, and I_{xx} is the second moment of area for the colony about the substrate. Assuming the cross-sectional area at the colony base is approximately elliptical, the second moment of area is²⁴:

$$I_{xx} = \frac{\pi d_{||}^3 d_{\perp}}{64} \quad (5)$$

$d_{\perp}$ is the width of the base perpendicular to water flow. Combining equations (4) and 5 yields:

$$\sigma_t = \frac{32 M}{\pi d_{||}^2 d_{\perp}} \quad (6)$$

Substituting equation (3) for M in equation (6), then rearranging, we obtain:

$$\sigma_t = \frac{16 \rho_w}{\pi d_{||}^2 d_{\perp}} U^2 \int_{y=0}^h y w(y) dy \quad (7)$$

Dislodgement occurs when this quantity equals or exceeds the tensile strength of the reef substrate, σ_s . Setting up this inequality, and rearranging, yields expression (1) in the main text.

Field test. The field test was conducted at Lizard Island in the northern Great Barrier Reef (GBR) lagoon, Australia. The study site, the southeastern reef, is typical of exposed reef platforms in this region; it is composed of steep reef slopes, shallow exposed crests and extensive reef flats. Three species of scleractinian reef

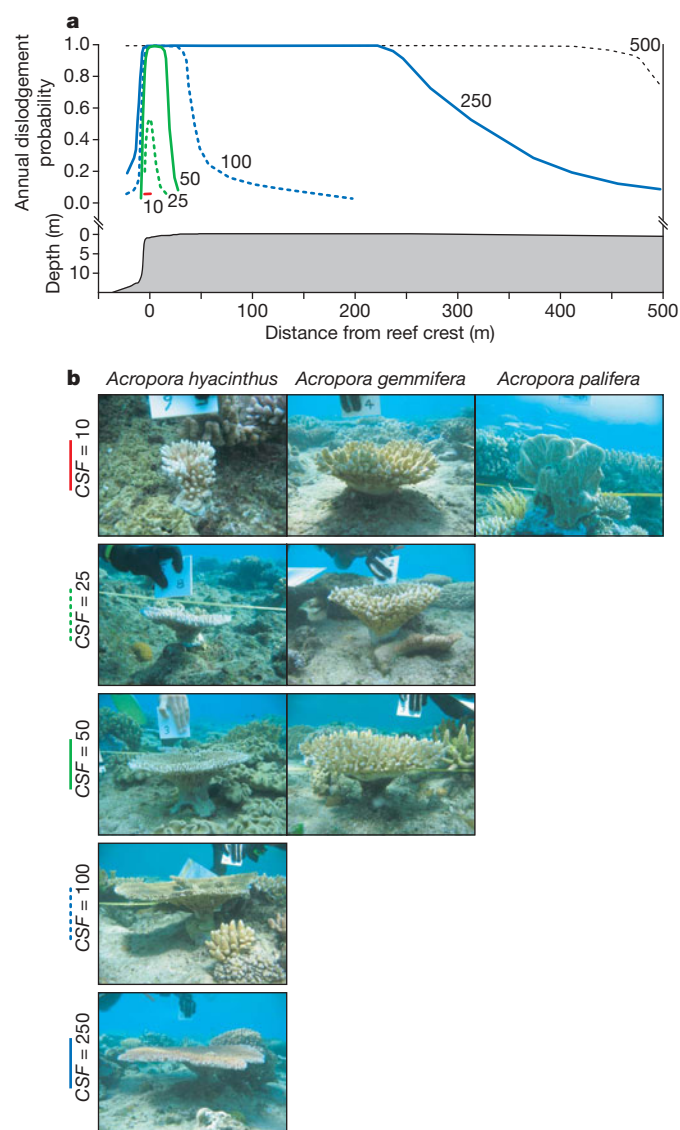


Figure 3 | Annual probabilities of colony dislodgement predicted from the 37-yr history of water velocity on the study reef. a, Yearly dislodgement estimates as a function of distance from the reef crest for six *CSF* values (10, 25, 50, 100, 250 and 500) that span the realistic range of values for a wave-exposed reef platform **b**, Images of colonies of the three species with *CSF* values approximately equal to five of those represented in the upper panel (no colonies at the study site were near *CSF* = 500).

coral were selected at these sites on the basis of two criteria: (1) individuals were abundant from the reef crest to 80 m across the flat towards the back of the reef, and (2) the species each had characteristic morphologies which were likely to exhibit different changes in *CSF* with increasing size (Fig. 3b). The first species, *Acropora hyacinthus*, is fast-growing and forms horizontal tables up to several square metres in size, and is typically attached to the substrate by a short central stalk. The second species, *Acropora gemmifera*, forms digitate colonies with a geometrically variable substrate attachment. The final species, *Acropora palifera* (subgenus *Isopora*), forms sturdy submassive mounds with the largest attachment area relative to colony size.

Eight belt transects (each 80 m long and 2 m wide) were laid perpendicular to the crest and parallel to the prevailing wave motion. The transects extended across the reef flat from the crest towards the reef back and all colonies of the three study species within the 160-m² transect areas were digitally photographed and their distance from the reef crest was recorded. For each colony, photographs were taken from above, as well as parallel and perpendicular to the reef crest (and thus the direction of prevailing wave climates⁹) to obtain the digital information required to calculate *CSFs*. For each photograph, the colony's outline, its basal points (where the outline of the colony coincided with the substrate) and the length of a 10 cm scale-plate were digitised and recorded as xy-coordinates. A computer program (available from the authors) measured the area encompassed by the colony outline, the width of the colony and the dimensions of the base from the digital images, then calculated the bending moment (per water velocity squared) numerically by calculating the product of each pixel within the colony outline and the height of that pixel from the colony base, and then summing this quantity over all pixels in the colony outline (equation (4)). Using these quantities, along with the basal width parallel to flow, $d_{||}$ (calculated by the same program), the *CSF* for each colony was calculated according to expression (1). Planar area, which gives an indication of the two-dimensional space occupied by a colony, was estimated from the colony photograph from above.

Received 18 July; accepted 9 October 2006.

- Liu, P. L.-F. et al. Observations of the international tsunami survey team in Sri Lanka. *Science* **308**, 1595 (2005).
- Baird, A. H. et al. Acehnese reefs in the wake of the Asian tsunami. *Curr. Biol.* **15**, 1926–1930 (2005).
- Hoyos, C. D., Agudelo, P. A. & Curry, J. A. Deconvolution of the factors contributing to the increase in global hurricane intensity. *Science* **312**, 94–97 (2006).
- Watson, R. T., Core Writing Team (eds) *3rd Assessment Report of the Intergovernmental Panel on Climate Change. Climate Change 2001: Synthesis Report* (www.ipcc.ch/pub/syngeng.htm) (2001).
- Emanuel, K. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436**, 686–688 (2005).
- Adger, W. N., Hughes, T. P., Folke, S. R., Carpenter, S. R. & Rockstrom, J. Social-ecological resilience to coastal disasters. *Science* **309**, 1036–1039 (2005).
- Done, T. J., Ogden, J. C., Wiebe, W. J. & Rosen, B. R. Diversity and ecosystem function of coral reefs. In: *Functional Roles of Biodiversity: a Global Perspective* (eds Mooney, J. H. et al) 393–429 (John Wiley, New York, 1996).
- Moberg, F. & Folke, C. Ecological goods and services of coral reef ecosystems. *Ecol. Econ.* **29**, 215–233 (1999).
- Madin, J. S., Black, K. P. & Connolly, S. R. Scaling water motion on coral reefs: from regional to organismal scales. *Coral Reefs* in the press, doi: 10.1007/s00338-006-0137-2 (August 2006).
- Massel, S. R. & Done, T. J. Effects of cyclone waves on massive coral assemblages on the Great Barrier Reef—meteorology, hydrodynamics and demography. *Coral Reefs* **12**, 153–166 (1993).
- Massel, S. R. *Fluid Mechanics for Marine Ecologists* 428–435 (Springer, Heidelberg, 1999).
- Rosen, B. R. The distribution of coral reefs. *Report Underwater Assoc.* **1**, 1–16 (1975).
- Done, T. J. Zonation: its nature and significance. In *Perspectives on coral reefs* (ed. Barnes, D. J.) 107–147 (Brian Clouston, Manuka, Australia, 1983).
- Connell, J. H. Diversity in tropical rainforests and coral reefs. *Science* **199**, 1302–1310 (1978).
- Hughes, T. P., Bellwood, D. R., Folke, C., Steneck, R. S. & Wilson, J. New paradigms for supporting the resilience of marine ecosystems. *Trends Ecol. Evol.* **20**, 380–386 (2005).
- Wainwright, S. A., Biggs, S. A., Currey, J. D. & Gosline, J. M. *Mechanical Design in Organisms* (Princeton University Press, Princeton, 1982).
- Denny, M. W. Predicting physical disturbance: mechanistic approaches to the study of survivorship on wave-swept shores. *Ecol. Monogr.* **65**, 371–418 (1995).
- Madin, J. S. Mechanical limitations of reef corals during hydrodynamic disturbances. *Coral Reefs* **24**, 630–635 (2005).
- Elsner, J. B. & Kara, A. B. *Hurricanes of the North Atlantic: Climate and Society* 339–354 (Oxford University Press, New York, 1999).
- Taylor, H. M. & Karlin, S. *An Introduction to Stochastic Modelling* Revised edn 36–38 (Academic Press, Boston (1994).
- Kench, P. S. Physical processes in an Indian Ocean atoll. *Coral Reefs* **17**, 155–168 (1998).
- Dollar, S. J. & Tribble, G. W. Recurrent storm disturbance and recovery: a long-term study of coral communities in Hawaii. *Coral Reefs* **12**, 223–233 (1993).
- Bellwood, D. R., Hughes, T. P., Folke, C. & Nyström, M. Confronting the coral reef crisis. *Nature* **429**, 827–833 (2004).
- Meriam, J. L. & Kraige, L. G. *Engineering Mechanics, Statics* 2nd edn (Wiley and Sons, New York, 1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Ridd, C. Berndt and especially T. Hughes for helpful discussions; and A. Baird, M. Berumen A. Dell, M. Dornelas, C. Dudgeon, E. Madin, M. Pratchett and the staff of the Lizard Island Research Station for help in the field. This work was supported by James Cook University and the Australian Research Council.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.S.M. (madin@nceas.ucsb.edu).

Human embryonic stem cell lines derived from single blastomeres

Irina Klimanskaya^{1*}, Young Chung^{1*}, Sandy Becker¹, Shi-Jiang Lu¹ & Robert Lanza¹

The derivation of human embryonic stem (hES) cells currently requires the destruction of *ex utero* embryos^{1–4}. A previous study in mice indicates that it might be possible to generate embryonic stem (ES) cells using a single-cell biopsy similar to that used in preimplantation genetic diagnosis (PGD), which does not interfere with the embryo's developmental potential⁵. By growing the single blastomere overnight, the resulting cells could be used for both genetic testing and stem cell derivation without affecting the clinical outcome of the procedure. Here we report a series of ten separate experiments demonstrating that hES cells can be derived from single blastomeres. In this proof-of-principle study, multiple biopsies were taken from each embryo using micromanipulation techniques and none of the biopsied embryos were allowed to develop in culture. Nineteen ES-cell-like outgrowths and two stable hES cell lines were obtained. The latter hES cell lines maintained undifferentiated proliferation for more than eight months, and showed normal karyotype and expression of markers of pluripotency, including Oct-4, SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, nanog and alkaline phosphatase. These cells retained the potential to form derivatives of all three embryonic germ layers both *in vitro* and in teratomas. The ability to create new stem cell lines and therapies without destroying embryos would address the ethical concerns of many, and allow the generation of matched tissue for children and siblings born from transferred PGD embryos.

A series of experiments was carried out to determine whether hES cells can be derived from single blastomeres (Table 1). Unused

embryos produced by *in vitro* fertilization (IVF) for clinical purposes were obtained with full informed consent and used in compliance with Advanced Cell Technology's ethics advisory board and institutional review board. Sixteen pronuclear- and multicell-stage embryos were thawed and cultured to the 8–10-cell stage in 20- μ l microdrops of Quinn's cleavage medium under paraffin oil (see Methods). Six of these embryos were grade I or II (symmetrical and even cell division with little or no cytoplasmic fragmentation), whereas the remaining ten embryos were grade III (variable fragmentation), using a standard scoring system⁶; embryos with blastomeres of unequal size and moderate-to-severe fragmentation (grades IV and V) were excluded from this study. The zona pellucida was disrupted and individual blastomeres mechanically separated from the denuded embryos using a micropipette and gentle tapping of the pipette holder. Multiple blastomere biopsies were obtained from each embryo to minimize the number of embryos used in this study. The separated blastomeres were cultured together in the same medium, arranged so as to avoid contact with each other using depressions created in the bottom of the plastic tissue culture plate, as described previously⁷.

The majority (58%) of the isolated blastomeres divided at least once, and approximately half of these (28 out of 53) formed vesicles/clumps that produced outgrowths within two days (Fig. 1). Previous experiments in mice⁵ indicated that cell co-culture is important for ES cell derivation from single blastomeres. However, the aggregation system used in these previous studies could not be employed,

Table 1 | Embryos used for embryonic stem-cell derivation

Experiment number	Embryo number	Grade of embryo	Number of blastomeres retrieved	Number of blastomeres remaining	Number of blastomeres that divided	Number of outgrowths	Number of ES-cell-like outgrowths	Number of ES cell lines established	Comments
1	1	III	5	3	3	1	0	0	without ES co-culture
1	2	III	5	3	1	0	0	0	without ES co-culture
2	3	III	6	4	3	0	0	0	without ES co-culture
3	4	I	7	1	4	3	3	1	ES co-culture
3	5	II	4	4	2	2	1	0	ES co-culture
4	6	III	7	3	6	1	1	0	ES co-culture
5	7	III	6	2	3	1	1	0	ES co-culture
5	8	III	6	2	4	2	2	0	ES co-culture
6	9	II	5	3	2	0	0	0	ES co-culture
6	10	I	7	1	5	5	4	1	ES co-culture
7	11	III	5	3	4	3	3	0	ES co-culture
7	12	III	6	2	3	1	0	0	ES co-culture
8	13	III	6	2	3	0	0	0	ES co-culture
9	14	III	4	4	3	3	2	0	ES co-culture
10	15	II	6	4	3	2	0	0	ES aggregation (different technique)
10	16	II	6	2	4	4	2	0	ES aggregation (different technique)
Total	16		91	43	53	28	19	2	

The zona pellucida was disrupted using either acidic Tyrode's solution (experiments 1–5) or multiple piezo-pulses (experiments 6–10). After biopsying, the embryos were not allowed to continue developing and were discarded.

¹Advanced Cell Technology, 381 Plantation Street, Worcester, Massachusetts 01605, USA.

*These authors contributed equally to this work.

because—unlike in the mouse—human blastomeres do not form tight aggregates with ES cells. Therefore, microdrops containing the blastomere-derived vesicles/clumps were merged with microdrops seeded with mitomycin-C-treated mouse embryonic fibroblasts (MEFs) and green fluorescent protein (GFP)-positive hES cells. When the outgrowths reached approximately 50–100 cells, they were mechanically passaged onto fresh feeders in microdrops containing hES cell medium. Although the initial outgrowths generally contained cells of different morphologies, over a period of several days we observed a number of fates: (1) cells resembling trophectoderm took over the culture, (2) cells that initially resembled ES cells differentiated within the culture, or (3) ES-cell-like cells continued undifferentiated proliferation. All of these outcomes are typical of derivation of ES cells from human embryos, especially when intact blastocysts are plated out without the removal of the trophectoderm using immunosurgery.

Routine passaging of putative hES cells was performed by mechanical dispersion of the colonies and transfer to fresh feeders every 2–3 days, until there were enough cells for trypsinization (after passage 10; see ref. 8). The colony morphology, growth rate, and procedures and culture media used were very similar to those of blastocyst-derived ES cells^{3,8,9}. Two of the six grade I/II embryos generated stable hES cell lines that showed normal karyotype (line MA01, 46,XX; line MA09, 46,XX; Fig. 2h) and molecular markers of pluripotency (Fig. 2a–g). Both lines stain for alkaline phosphatase, and express octamer binding protein 4 (Oct-4), stage-specific embryonic antigen (SSEA)-3, SSEA-4, tumour-rejection antigen (TRA)-1-60, and TRA-1-81. Karyotype analysis ruled out fusion (both new lines were female and the WA01 hES cells used for co-culture were male; Fig. 2h), and microsatellite analysis confirmed that MA01 and MA09 were genetically distinct from other hES cell lines grown in our laboratory (Fig. 3d). Polymerase chain reaction (PCR) analysis further confirmed this, indicating the absence of GFP and Y chromosome gene sequences in both of the blastomere-derived hES cell lines (Fig. 3a–c).

When the hES cell cultures were allowed to overgrow or form embryoid bodies, they readily differentiated into cells of all three germ layers, as evidenced by immunostaining with antibodies to α -fetoprotein (primitive endoderm; Fig. 4b), smooth-muscle actin (mesoderm; Fig. 4c) and β III tubulin (ectoderm; Fig. 4d). The single-blastomere-derived hES cells could also be differentiated *in vitro* into cells of specific therapeutic interest, including endothelial cells that, after replating on Matrigel, formed typical capillary/vascular-like structures (Fig. 4e) that expressed high levels of von Willebrand

factor and took up acetylated low-density lipoprotein (Fig. 4f). Retinal pigment epithelium (RPE) clusters also appeared in adherent hES cell cultures and in embryoid bodies, and were used to establish, as described previously¹⁰, passable RPE cell lines that displayed a pigmented phenotype and typical 'cobblestone' morphology (Fig. 4g), bestrophin immunostaining (Fig. 4h), and expressed bestrophin, *RPE65*, *CRALBP* and *PEDF* as shown by PCR with reverse transcription (RT-PCR; Fig. 4i). To demonstrate the pluripotency of the putative ES cells derived in this study, cells were injected into NOD-SCID mice forming teratomas containing tissues from all three germ layers including neural rosettes (ectoderm), haematopoietic cells (mesoderm), and liver, respiratory and intestinal

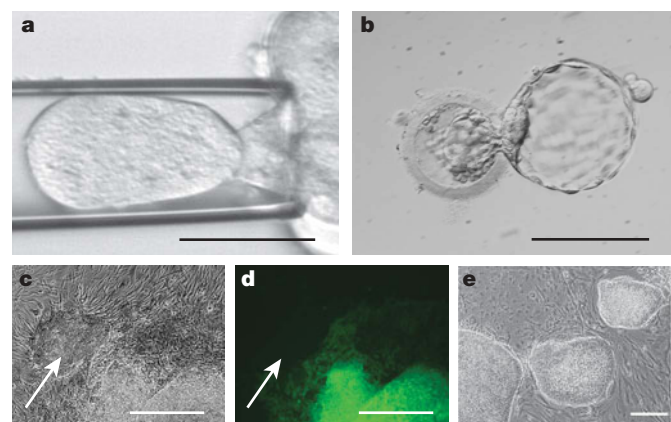


Figure 1 | Derivation of hES cells from single blastomeres. **a**, Biopsy of a single blastomere. **b**, Development of a blastomere-biopsied embryo into a hatching blastocyst. **c**, **d**, Blastomere-derived outgrowth (arrows) close to a colony of GFP-positive hES cells. The same field of cells is shown under phase contrast (**c**) and fluorescence (**d**). **e**, Morphology of blastomere-derived hES cell colonies. Scale bars: 50 μ m for **a**; 200 μ m for **b–e**.

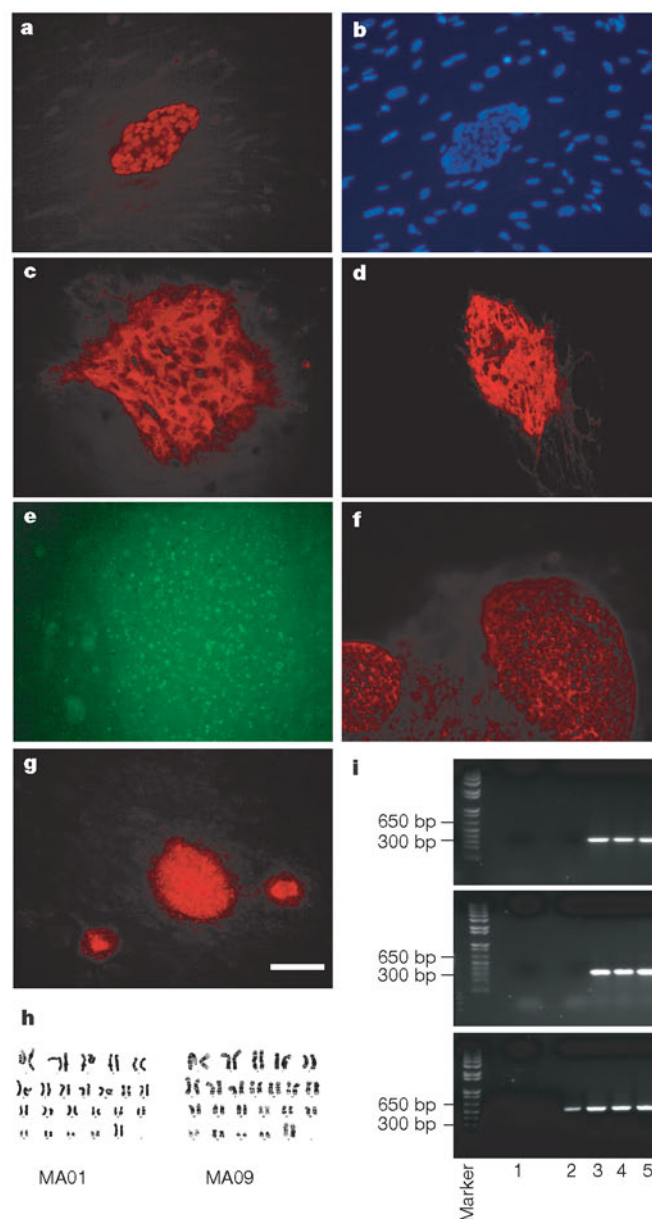


Figure 2 | Characterization of hES cells from single blastomeres. **a–g**, Staining for markers of pluripotency, showing Oct-4 (**a**) and corresponding DAPI staining (**b**), TRA-1-60 (**c**), TRA-1-81 (**d**), SSEA-3 (**e**), SSEA-4 (**f**) and alkaline phosphatase (**g**). Scale bar in **g** indicates 50 μ m for panel **e** and 100 μ m for panels **a–d** and **f** and **g**. **h**, Representative chromosome spreads of the two single-blastomere-derived hES cell lines. **i**, RT-PCR analysis of the expression of markers of pluripotency in single-blastomere-derived hES cell lines. Top panel, *Oct-4*; centre panel, *nanog*; bottom panel, *GAPDH*. Lane 1, no template; lane 2, negative control (MEFs); lane 3, MA01; lane 4, MA09; lane 5, WA01.

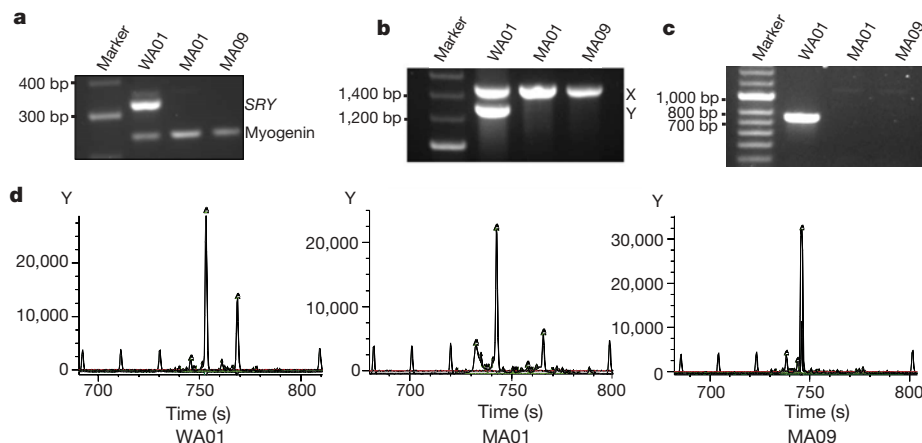


Figure 3 | Microsatellite and PCR analysis of single-blastomere-derived hES cells versus the cell line (WA01) used for co-culture.

a, b, SRY (**a**) and amelogenin (**b**) genes (cropped image). **c,** GFP sequences (cropped image). **d,** Microsatellite analysis. Full-length gel images are presented in Supplementary Fig. 1.

epithelia (endoderm), among others (Fig. 4a). The presence of all three germ layers was also confirmed by immunostaining (Fig. 4a, insets). Although only two of the six (33%) grade I/II embryos (or 2 out of 35 blastomeres; 2 out of 91 blastomeres including grade III–V embryos) used in the current study generated hES cell lines, this success rate is similar to that reported by other

investigators using conventional derivation methods^{2,3,11,12}. The grade of the leftover embryos proved crucial, and we believe the success rate can be further increased by optimizing conditions at the earliest stages of blastomere outgrowth. The two stable hES cell lines were generated from the highest quality (grade I) embryos in the study (Table 1). Because we had very few grade I embryos,

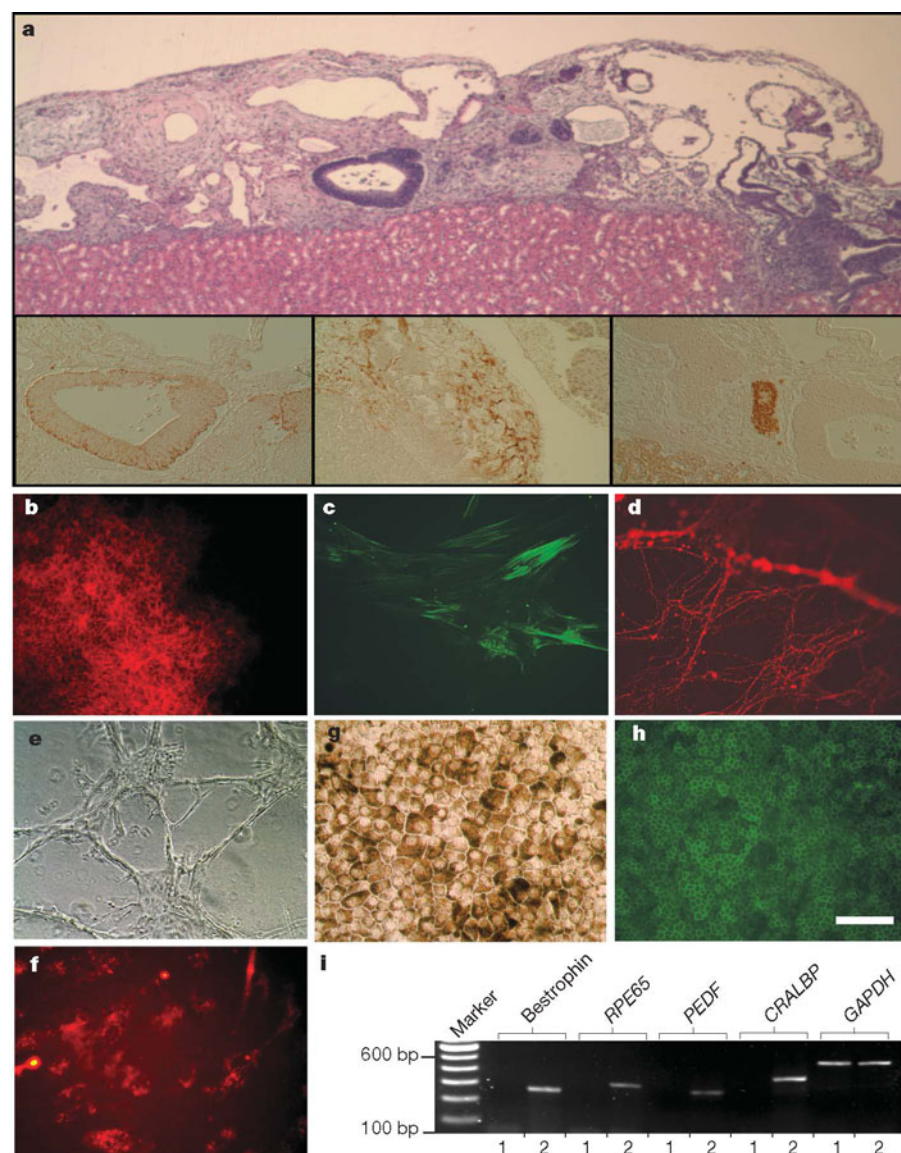


Figure 4 | In vitro differentiation of single-blastomere-derived hES cells into all three germ layers.

a, Teratoma formation after transplantation of hES cells under the kidney capsule of NOD-SCID mice for seven weeks. Insets show enlargements of adjacent sections stained for molecular markers to confirm the presence of all three germ layers: left inset, neural tissue stained for nestin (ectoderm); centre inset, smooth-muscle actin (mesoderm); right inset, intestine stained for CDX2 (endoderm). Original magnifications for **a** are $\times 40$ (top), $\times 200$ (bottom left and right) and $\times 100$ (bottom middle). **b–d,** Immunofluorescence analysis of the molecular markers of primitive endoderm (**b**, α -fetoprotein), mesoderm (**c**, smooth-muscle actin) and ectoderm (**d**, β III tubulin). **e–i,** In vitro differentiation of single-blastomere-derived hES cells into cells of specific therapeutic interest. **e,** Endothelial cells plated on Matrigel showing formation of typical capillary-tube-like structures. **f,** Acetylated low-density lipoprotein uptake by endothelial cells. **g, h,** Retinal pigment epithelium (RPE) showing a pigmented phenotype and typical 'cobblestone' morphology (**g**) and bestrophin staining (**h**). Scale bars, 100 μ m for **a–h**. **i,** RT-PCR analysis for markers of RPE in single-blastomere-derived RPE (cropped image). For each gene: lane 1, negative control (undifferentiated hES cells, line WA09); lane 2, RPE from hES cells. A full-length gel image is presented in Supplementary Fig. 1.

almost every blastomere was removed and cultured in the same dish to maximize the chances of success. It is possible that healthier embryos are more likely to generate a robust growth response, although it is unclear to what extent the interaction and commitment of the biopsied blastomere affected this process.

According to the Food and Drug Administration (FDA) regulations, the presence of animal cells in culture systems is permissible under current xenotransplantation guidelines¹³. We are in the process of adapting both of the blastomere-derived hES cell lines (which were derived on mouse feeders) to animal-free conditions under current good manufacturing practice (cGMP) to comply with these requirements. However, it may be desirable to derive future lines using human feeder cells, or feeder-free systems such as those that have been described previously by our group⁹ and others^{14–16}.

Additional studies will also be necessary to determine whether blastomere-derived hES cell lines differ from conventional hES cell lines in their ability to form functional differentiated cell types. Blastomere-derived lines MA01 and MA09 were observed to differentiate, at the very least, in the same manner and at the same rate as all other hES cell lines studied in our laboratory. Furthermore, it seems that they may more readily differentiate into certain vital cell types, although it is unclear whether this is related to the derivation technique or number of passages by trypsinization. For instance, neural progenitors were readily generated without the need for embryoid-body intermediates, stromal feeder layers, or low-density passaging. When transferred to a laminin-coated substrate and maintained in defined medium containing laminin and basic fibroblast growth factor (FGF2), they began to express neural and neural progenitor markers, including nestin, β III tubulin and Pax6, at the second passage. MA01 hES cells also formed haematopoietic colony forming units (CFUs) 3–5 times more efficiently than WA01-GFP cells and 5–10 times more efficiently than WA09 cells. (NIH-registered WA01 and WA09 hES cell lines were formerly known as H1 and H9, respectively²). MA09 hES cells showed similar potential as WA09 cells for haematopoietic differentiation, but demonstrated a higher capability to differentiate towards the endothelial lineage as compared with both WA01-GFP and WA09 cells (data not shown).

Concerns have been raised as to whether individual eight-cell-stage blastomeres, such as those used in the current study, are totipotent and could potentially generate a human being. A recent report shows the localization of cell fate determinants (*Cdx2*) in the late-dividing blastomeres of the two-cell-stage mouse embryo¹⁷. Other studies show that at the four-cell stage, blastomeres have different developmental properties¹⁸, and that individual human blastomeres have differential *Oct-4* expression that seems to direct cells towards intermediate cell mass or trophoctoderm lineages¹⁹. Notably, individual morula (8–16-cell)-stage blastomeres have never been shown to have the intrinsic capacity to generate a complete organism in any mammalian species.

Previous attempts to induce isolated human blastomeres to proliferate into pluripotent stem cell lines have failed^{20,21}. One study²⁰ observed proliferation of blastomeres (in the range of 1–8 cells per blastomere) from cleavage-stage human embryos *in vitro*, although in all cases differential labelling indicated that they had generated trophoctodermal cells. Here we demonstrate that single blastomeres can be used to establish hES cell lines. This technique is modelled on PGD, an approach for single-cell biopsy using mechanical micromanipulation that does not interfere with the developmental capacity of the parent embryo. Figure 1b shows an image from a control experiment which confirmed that the developmental rate (75%) of eight single-cell biopsied 8-cell-stage embryos to the blastocyst stage is similar to that for four non-biopsied embryos. Blastomeres grown overnight could be used for both genetic testing and stem cell generation, thereby allowing selection for a day-5 blastocyst transfer. Numerous reports suggest that neither the survival rate nor the subsequent development and chances of implantation differ between intact human embryos at the blastocyst stage and those following

blastomere biopsy for PGD^{22–25}. However, until remaining doubts about safety are resolved, we do not recommend this procedure be applied outside the context of PGD. Blastomere-derived hES cells could be of great potential benefit for medical research, as well as for children and siblings born from transferred PGD embryos.

METHODS

Derivation of hES cells. Human embryos were thawed and cultured until the 8–10-cell stage at 37 °C in 20- μ l drops of Quinn's cleavage medium under paraffin oil in a highly humidified incubator with 5.5% CO₂ in air. The zona pellucida was disrupted using either acidic Tyrode's solution or multiple piezo-pulses, and individual blastomeres were mechanically separated from the denuded embryos by holding the cell with a micropipette and gently tapping the pipette holder. The separated blastomeres were cultured together in the same medium, arranged so as to avoid contact with each other using depressions created in the bottom of the plastic tissue culture plate as described previously⁷. During this process, sets of microdrops were prepared, consisting of a 50- μ l drop of Quinn's blastocyst medium (Cooper Surgical) supplemented with 5 μ g ml⁻¹ human placental laminin and human plasma fibronectin (Sigma), surrounded by several microdrops of hES cell culture medium⁸ containing GFP-labelled hES cells growing on a MEF feeder layer. The microdrops containing the blastomere-derived aggregates were merged with one or two of the surrounding microdrops by scraping the bottom of the plate between the drops with a glass capillary, and fresh MEFs added to the drops in 24 h. After the formation of initial outgrowths, approximately half of the medium was changed every other day until the outgrowths reached approximately 50–100 cells. They were then mechanically passaged onto fresh MEF feeder layers in hES cell culture medium, which was changed every 1–2 days. The colonies were passaged by mechanical dispersion until enough cells were produced to initiate adaptation to trypsin. Thereafter, they were cultured as described previously⁸. We will provide a detailed protocol elsewhere for the derivation and maintenance of these cells (*Nature Protocols*, manuscript in preparation).

Immunostaining. Cells were fixed with 2% (w/v) paraformaldehyde, permeabilized with 0.1% (v/v) NP-40, and blocked with 10% (v/v) goat serum, 10% (v/v) donkey serum. Incubation with primary antibodies was carried out overnight at 4 °C. The antibodies used in this study were anti-Oct-4 (Santa Cruz Biotechnology), anti-SSEA-3 and anti-SSEA-4 (developed by Solter and Knowles and obtained from the Developmental Studies Hybridoma Bank, University of Iowa), anti-TRA-1-60 and anti-TRA-1-81 (Chemicon), anti- β -III-tubulin (Covance), anti- α -fetoprotein (DAKO) and anti-smooth-muscle-actin (DAKO). After washing in PBS containing 0.1% (v/v) Tween-20 (PBST), fluorescently labelled or biotinylated secondary antibodies were added for 1 h; some samples were subsequently incubated for 15 min with fluorescently labelled streptavidin. After additional washing in PBST, specimens were mounted using Vectashield with DAPI (4,6-diamidino-2-phenylindole; Vector Laboratories) and observed using a fluorescent microscope.

Karyotyping. The cells were passaged onto gelatin in ES cell culture medium, which was replaced the day before harvest when the cells were approximately 50% confluent. Colcemid was added to the culture at a concentration of 0.12 μ g ml⁻¹ for 40 min, then the cells were rinsed twice with PBS, trypsinized and centrifuged in DMEM with 10% (v/v) fetal bovine serum. KCl (0.075 M) was added to the pellet, and the cells were incubated for 10 min at 37 °C, centrifuged and fixed with 3:1 methanol:acetic acid for 10 min, then centrifuged again and suspended in this fixative. Cytogenetic analysis was performed on metaphase cells using G-banding on ten cells.

Microsatellite and PCR analyses. For microsatellite analysis, genomic DNA (MA01, passage 18; MA09, passage 23) was extracted with a DNeasy Tissue kit (Qiagen). Conventional PCR reactions were performed with 100 ng gDNA, AmpliTaq Gold polymerase (ABI), and primer pairs specific for FES/FPS, vWA31, D2S417, D10S526 and D5S592 genomic microsatellite sequences (Coriell). Single primers in each pair were end-labelled with a 6-Fam fluorescent label. After incubation for 10 min at 94 °C to activate the polymerase, amplification was performed with 30 cycles of 94 °C for 45 s, 56 °C for 60 s, and 72 °C for 60 s. Labelled amplicons were separated and sized using an ABI 3730 sequencer (ABI). For amplification of enhanced GFP, amelogenin and SRY genes, gDNA (MA01, passage 40; MA09, passage 44) was isolated using a QIAamp DNA Mini kit (Qiagen), and 200 ng gDNA per reaction in 50 μ l was used for GFP, amelogenin and SRY amplification. The primers used for GFP were 5'-TTGAATTGCGCCACCATGGTGAGC-3' (forward) and 5'-TTGAATTCTTAC-TTGTACAGCTCGTCC-3' (reverse), and PCR reactions were performed as described previously⁵. For sex determination, both the amelogenin and SRY genes were amplified by PCR as described previously^{26,27}. The primers used for the amelogenin gene were 5'-CTCATCCTGGGCACCCCTGTTATATC-3'

(forward) and 5'-GGTACCACTTCAAAGGGTAAGCAC-3' (reverse), which generated a fragment of 1,310 base pairs (bp) for the Y chromosome and a fragment of 1,490 bp for the X chromosome. For the Y-chromosome-specific SRY gene, the primers used were 5'-GATCAAGCAAGCAGCTGGGATACCAAGTG-3' (forward) and 5'-CTGTAGCGGTCCCGTTGCTGCGGTG-3' (reverse), which amplified a DNA fragment of 330 bp. As a control for the PCR reactions, myogenin primers (5'-TCACGGTGGAGGATATGTCT-3' (forward) and 5'-GAGTCAGCTAAATCCCTCG-3' (reverse)) were included in SRY PCR reactions, which generated a fragment of 245 bp. PCR products were separated on an agarose gel and visualized by ethidium bromide staining.

Total RNA was isolated using the RNeasy Mini kit (Qiagen); for RPE cells, a Trizol purification step was included. RT-PCR was performed using the Qiagen One Step RT-PCR kit for hES cell markers using the following conditions: 50 °C for 30 min for reverse transcription; 95 °C for 15 min for inactivation of the reverse transcriptase and activation of the polymerase; 30 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min; and 72 °C for 10 min for final extension. The primers used were for *Oct-4* (5'-GAAGGTATTCAGCCAAACGAC-3' (forward) and 5'-GTTACAGAACCACACTCGGA-3' (reverse); 315 bp), *nanog* (5'-TGCAATGTCTTCTGCTGAGAT-3' (forward) and 5'-GTTCAGGATGTTGGAGAGTTC-3' (reverse); 285 bp) and *GAPDH* (5'-GTCCATGCCATCACTGCCA-3' (forward) and 5'-TTACTCCTTGGAGGCCATG-3' (reverse); 513 bp). For RPE markers, complementary DNA was made using the SuperScript III CellDirect Synthesis system (Invitrogen). RT-PCR was performed using Platinum Blue PCR SuperMix (Invitrogen). After incubation at 94 °C for 3 min, 30 cycles of 94 °C for 35 s, 55 °C for 35 s, and 72 °C for 50 s were performed, followed by a final extension at 72 °C for 10 min. An undifferentiated hES cell line WA09 was used as a negative control. The primers used were *RPE65* (5'-ATGGACTTGGCTTGAATCACTT-3' (forward) and 5'-GAACAGTCCATGAAGGTGACA-3' (reverse); 285 bp), *bestrophin* (5'-ATCAGAGGCCAGGCTACTACAG-3' (forward) and 5'-TCCACAGTTTCTCTCACTT-3' (reverse); 235 bp), *RALBP* (*CRALBP*) (5'-AAATCAATGGCTTCTGCATCACTT-3' (forward) and 5'-CCAAAGAGCTGCTCAGCAAC-3' (reverse); 340 bp), *PEDF* (5'-AAGCTGAGTTATGAAGGCGAAG-3' (forward) and 5'-TCTGTGTCCCTCAGTACGAAGA-3' (reverse); 255 bp) and *GAPDH* (5'-CGATGCTGCGCTGAGTAC-3' (forward) and 5'-GTTTCTTACTCCTTGGAGGC-3' (reverse); 465 bp).

Teratomas. Small clumps of 50–100 cells were mechanically removed from the culture and transplanted under the kidney capsules of 6–8-week-old NOD-SCID mice under anaesthesia. Between 18–57 days after transplantation, the kidneys were removed, fixed with 4% (w/v) paraformaldehyde overnight, washed for 24 h in PBS, embedded in paraffin, sectioned and analysed for the presence of the derivatives of the three germ layers.

Received 12 April; accepted 8 August 2006.

Published online 23 August; amended on 23 November 2006, changes are detailed in an accompanying Addendum (doi:10.1038/nature05366).

1. Lanza, R. *et al.* (eds) *Essentials of Stem Cell Biology* (Elsevier/Academic, San Diego, 2006).
2. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
3. Cowan, C. A. *et al.* Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350**, 1353–1356 (2004).
4. Strelchenko, N. *et al.* Morula-derived human embryonic stem cells. *Reprod. BioMed.* **9**, 623–629 (2004).
5. Chung, Y. *et al.* Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. *Nature* **439**, 216–219 (2006).
6. Veeck, L. L. *et al.* *An Atlas of Human Gametes and Conceptuses* (Parthenon, New York, 1999).
7. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. (eds) *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2002).
8. Klimanskaya, I. & McMahon, J. in *Handbook of Stem Cells* Vol. 1 *Embryonic Stem Cells* (eds Lanza, R. *et al.*) 437–449 (Elsevier/Academic, San Diego, 2004).

9. Klimanskaya, I. *et al.* Human embryonic stem cells derived without feeder cells. *Lancet* **361**, 1496–1501 (2005).
10. Klimanskaya, I. *et al.* Derivation and comparative assessment of retinal pigment epithelium from human embryonic stem cells using transcriptomics. *Clon. Stem Cells* **6**, 217–245 (2004).
11. Lanzendorf, S. E. *et al.* Use of human gametes obtained from anonymous donors for the production of human embryonic stem cell lines. *Fertil. Steril.* **76**, 132–137 (2001).
12. Mitalipova, M. *et al.* Human embryonic stem cell lines derived from discarded embryos. *Stem Cells* **21**, 521–526 (2003).
13. Weber, D. J. Considerations for clinical uses of therapies derived from stem cells. In *Methods in Enzymology* Vol. 420 *Stem Cell Tools and Therapeutic Applications* (eds Klimanskaya, I. & Lanza, R.) (Elsevier, New York, in the press).
14. Richards, M., Fong, C. Y., Chan, W. K., Wong, P. C. & Bongso, A. Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells. *Nature Biotechnol.* **20**, 933–936 (2002).
15. Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nature Biotechnol.* **24**, 185–187 (2006).
16. Young, H., Xu, C. & Carpenter, M. K. in *Essentials of Stem Cell Biology* (eds Lanza, R. *et al.*) 317–324 (Elsevier/Academic, San Diego, 2006).
17. Deb, K., Sivaguru, M., Yong, H. Y. & Roberts, R. M. *Cdx2* gene expression and trophectoderm lineage specification in mouse embryos. *Science* **311**, 992–996 (2006).
18. Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S. & Zernicka-Goetz, Z. Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479–490 (2005).
19. Hansis, C., Tang, Y., Grifo, J. A. & Krey, L. C. Analysis of *Oct-4* expression and ploidy in individual human blastomeres. *Mol. Hum. Reprod.* **7**, 155–161 (2001).
20. Geber, S., Winston, R. M. L. & Handyside, A. H. Proliferation of blastomeres from cleavage stage human embryos *in vitro*: an alternative to blastocyst biopsy for preimplantation diagnosis. *Hum. Reprod.* **10**, 1492–1496 (1995).
21. Sills, E. S., Takeuchi, T., Tanaka, N., Queenie, V. N. & Palermo, G. D. Identification and isolation of embryonic stem cells in reproductive endocrinology: theoretical protocols for conservation of human embryos derived from *in vitro* fertilization. *Theor. Biol. Med. Model.* **2**, 25 (2005).
22. Handyside, A. H. *et al.* Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* **344**, 768–770 (1990).
23. Magli, M. C. *et al.* Cryopreservation of biopsied embryos at the blastocyst stage. *Hum. Reprod.* **21** (10), 2656–2660 (2006).
24. Hardy, K., Martin, K. L., Leese, H., Winston, R. M. & Handyside, A. Human preimplantation development *in vitro* is not adversely affected by biopsy at the 8-cell stage. *Hum. Reprod.* **5**, 708–714 (1990).
25. Staessen, C. *et al.* Comparison of blastocyst transfer with or without preimplantation genetic diagnosis for aneuploidy screening in couples with advanced maternal age: a prospective randomized controlled trial. *Hum. Reprod.* **19**, 2849–2858 (2004).
26. Fredsted, T. & Villesen, P. Fast and reliable sexing of prosimian and human DNA. *Am. J. Pathol.* **64**, 345–350 (2004).
27. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117–121 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Li and A. Feng for their assistance with hES cell culture and differentiation, S. Agarwal, J. Shepard and T. Lin for RT-PCR, J. Johnson and L. Meisner for karyotype analysis, and J. Staruk and L. McGill for histological analysis.

Author Information The hES cell lines generated in this study will be made available to investigators under a material transfer agreement (an application has been submitted for deposition of the hES cell lines into the UK Stem Cell Bank). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.L. (rlanza@advancedcell.com).

LETTERS

Endogenous neurosteroids regulate GABA_A receptors through two discrete transmembrane sites

Alastair M. Hosie¹, Megan E. Wilkins¹, Helena M. A. da Silva¹ & Trevor G. Smart¹

Inhibitory neurotransmission mediated by GABA_A receptors can be modulated by the endogenous neurosteroids, allopregnanolone and tetrahydro-deoxycorticosterone¹. Neurosteroids are synthesized *de novo* in the brain during stress², pregnancy³ and after ethanol consumption⁴, and disrupted steroid regulation of GABAergic transmission is strongly implicated in several debilitating conditions such as panic disorder, major depression, schizophrenia, alcohol dependence and catamenial epilepsy^{3,5–8}. Determining how neurosteroids interact with the GABA_A receptor is a prerequisite for understanding their physiological and pathophysiological roles in the brain. Here we identify two discrete binding sites in the receptor's transmembrane domains that mediate the potentiating and direct activation effects of neurosteroids. They potentiate GABA responses from a cavity formed by the α -subunit transmembrane domains, whereas direct receptor activation is initiated by interfacial residues between α and β subunits and is enhanced by steroid binding to the potentiation site. Thus, significant receptor activation by neurosteroids relies on occupancy of both the activation and potentiation sites. These sites are highly conserved throughout the GABA_A receptor family, and their identification provides a unique opportunity for the development of new therapeutic, neurosteroid-based ligands and transgenic disease models of neurosteroid dysfunction.

Endogenous neurosteroids are potent modulators of all major GABA_A receptor isoforms^{1,9}. At low nanomolar concentrations, observed during stress¹⁰, alcohol intoxication⁴ and oestrus¹¹, they potentiate GABA currents^{12–14}, whereas at submicromolar-to-micromolar concentrations, which occur during parturition³, they directly activate the receptor¹⁵. This profile indicates that two distinct neurosteroid binding sites may exist on the GABA_A receptor, but their location has remained enigmatic¹. Previous studies have proposed that neurosteroids target the GABA_A receptors' transmembrane (M) domains^{16–18}. To determine how important these are, we exploited the *Drosophila melanogaster* RDL (resistance to dieldrin) GABA receptor, which has very low sensitivity to potentiation and lacks direct activation by micromolar concentrations of neurosteroids¹⁹. We replaced M1 through to the end of M2 in the murine $\alpha 1$ and $\beta 2$ subunits with the corresponding sequence from the RDL subunit, forming the chimaeras αR and βR , respectively (Fig. 1a), and established their pharmacology as heterologously expressed $\alpha\beta\gamma$ subunit receptors. Potentiation and direct activation by tetrahydro-deoxycorticosterone (THDOC) and allopregnanolone (ALLOP) was abolished on receptors incorporating αR (Fig. 1b, and Supplementary Fig. 1), whereas receptors containing βR were indistinguishable from the wild type (data not shown). Thus, residues in the $\alpha 1$ -subunit M1 and/or M2 domains are critical for neurosteroid action.

To regulate GABA_A receptors, neurosteroids require a C3 α hydroxyl group on their A-ring and a C20 ketone in the D-ring²⁰ (Fig. 1c). These groups are important for the binding of other

steroids to a wide variety of proteins by means of hydrogen-bonding with polar or charged residues^{21,22}. In particular, several polar residues in the $\alpha 1$ subunit M1–M2 domains are replaced by hydrophobic residues in αR (T229I, T236I, Q241W, T264V, S269M, N274A and S275A), which could account for the reduced neurosteroid sensitivity of $\alpha R\beta 2\gamma 2$ receptors (Fig. 1a). Sequential replacement of these polar residues in the $\alpha 1$ subunit with the homologous RDL variants identified Thr 236 and Gln 241 as important determinants of steroid action. Mutating Thr 236 to Ile markedly reduced receptor activation by THDOC and ALLOP without affecting their potentiation of responses to GABA (Fig. 1d, and Supplementary Table 1). By contrast, Q241W ablated the potentiation of GABA currents and markedly reduced the apparent neurosteroid agonist efficacy without affecting the concentration causing half-maximal receptor activation (Fig. 1e). This mutation did not affect the potentiation mediated by diazepam, pentobarbitone or mefenamic acid, which bind to separate sites on GABA_A receptors (Fig. 1f). The distinctive functions of Thr 236 and Gln 241 in activation and potentiation by neurosteroids indicate that they may contribute to two discrete binding sites.

Using the transmembrane region of the nicotinic acetylcholine receptor²³, we created a homology model to determine where in the GABA_A receptor M domains Thr 236 and Gln 241 are located (Figs 2a and 3a, and Supplementary Fig. 4a). Threonine 236 lies on the receptor surface, close to the β/α subunit interface (the same interface in which GABA binds but to the amino-terminal domain) and would be accessible to hydrophobic neurosteroid molecules in the membrane (Fig. 3a). By contrast, Gln 241 lies at the base of a water-filled cavity between the α subunit's M1–M4 domains (Fig. 2a, and Supplementary Fig. 4a). After receptor activation, the depth and volume of this cavity seems to increase^{24–26}, which could allow neurosteroids to bind to Gln 241 and stabilize the receptor in an active state. The relative locations of Thr 236 and Gln 241 on opposite faces of the M1 helix strongly suggest that they will contribute to steroid binding at two distinct sites. We therefore determined how the hydrogen-bonding capacity of these residues affected neurosteroid potentiation and activation on GABA_A receptors.

Both the Q241H and Q241N variants supported wild-type levels of potentiation by ALLOP and THDOC, whereas their potencies were slightly reduced by Q241S and Q241T mutations (Fig. 2b, and Supplementary Fig. 1b). The level of potentiation was closely correlated with the substituent's abilities to act as hydrogen-bond acceptors (by providing a lone pair of electrons), which is how histidine residues form hydrogen bonds at pH 7.4. However, hydrophobic replacement with isoleucine or leucine abolished neurosteroid potentiation, while minimal potentiation was retained with methionine (10 μ M or more; Fig. 2c, and Supplementary Fig. 1b). Significantly, potentiation was also abolished with tryptophan, which acts only as a hydrogen-bond donor. Because the neurosteroid's C20 ketone can only participate as a hydrogen-bond acceptor, these data

¹Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK.

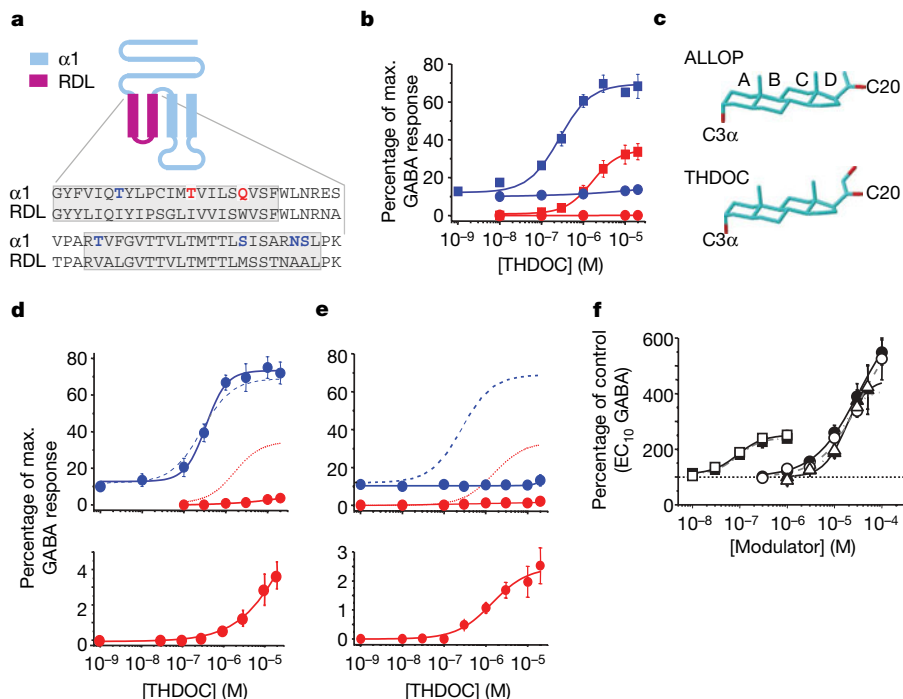


Figure 1 | Neurosteroid activity is determined by α -subunit M1 domain residues. **a**, The α R chimaera formed between α 1 and RDL subunits. Polar residues present only in α 1 (blue) are in bold, with Thr 236 and Gln 241 (red) highlighted. The transmembrane domains are boxed.

b, THDOC concentration–response curves for direct activation (red) and for potentiation (blue) of EC₁₀ (concentration causing 10% of maximal response) GABA currents expressed as a percentage of the maximum GABA response, at wild-type α 1 β 2 γ 2 (squares) and chimaeric α R β 2 γ 2 (circles) receptors. **c**, Structures of ALLOP and THDOC. **d**, **e**, THDOC concentration–response curves for α 1T236I β 2 γ 2 (**d**) and α 1Q241W β 2 γ 2 (**e**) mutants. Blue points, potentiation; red points, activation. Top: the wild-type activation (dotted) and potentiation (dashed) curves are taken from **b**. Bottom: reduced potency of activation by T236I (**d**) and reduced efficacy of activation in Q241W (**e**). **f**, Concentration–response curves for the modulation of EC₁₀ GABA currents by diazepam (squares), pentobarbitone (triangles) and mefenamic acid (circles). Filled symbols, wild type; open symbols, Q241W. Error bars represent s.e.m.

indicate that the C3 α hydroxyl group donates a hydrogen bond to Gln 241.

By using our homology model to locate other polar residues in this cavity that might engage with the neurosteroids' C20 ketone group, we identified Asn 407 and Tyr 410 in the M4 domain of the α 1 subunit. Both residues could donate a hydrogen bond to the C20 ketone and are separated from Gln 241 by about 15–18 Å (Fig. 2a), an appropriate distance to accommodate a neurosteroid molecule^{21,22}. The hydrophobic substitutions N407A and Y410F reduced neurosteroid potency although not to the same extent as the mutation of Gln 241 (3–10-fold opposed to more than 100-fold; Fig. 2d, and Supplementary Table 1). Furthermore, N407V and N407D, neither of which can act as a hydrogen-bond donor, caused similar reductions in neurosteroid potency (Supplementary Table 1). The mutation N407A caused a greater reduction in the potency of THDOC (about tenfold) than in that of ALLOP (about threefold), whereas Y410F only reduced the potency of THDOC (Fig. 2d, and Supplementary Fig. 1c). This is consistent with their coordination of oxygen moieties on the neurosteroid D-ring, the only region where ALLOP and THDOC vary structurally (Fig. 1c). A combined mutation of N407A and Y410F further reduced THDOC potency (Fig. 2d), whereas mutation of other polar residues in M3 and M4 (Y293F, S298A, N407A, Y410F, T413I and Y414F) had no effect. Potentiation of GABA responses by diazepam remained unaffected by all these mutations (data not shown). Q242, N407 and Y410 are therefore ideally located for a neurosteroid-binding site that potentiates receptor function.

We next determined whether α Thr 236 contributes to a second binding site associated with receptor activation. Replacing α Thr 236 with non-hydrogen-bonding isoleucine or valine markedly reduced the agonist potency of ALLOP and THDOC (Fig. 3b). By contrast, ALLOP and THDOC had similar potencies but lower apparent efficacies on α T236S β γ than on wild-type receptors (Fig. 3b), indicating that the presence and orientation of a hydroxyl group at position 236 might be critical for wild-type neurosteroid activation (Fig. 3b). If α Thr 236 contributes to a neurosteroid-binding site, it should lie about 15 Å from a second surface-exposed residue capable of hydrogen-bonding with the neurosteroid molecule. In our model, Tyr 284 in the β 2 subunit's M3 domain fulfilled these criteria (Fig. 3a). Consistent with a role in binding, the replace-

ment of β Tyr 284 with phenylalanine (removing the hydrogen-bonding hydroxyl group) markedly reduced the potencies of both THDOC and ALLOP as direct agonists (Fig. 3b) but had no effect on their potentiation of GABA currents (Supplementary Fig. 1i). The effects of mutating Thr 236 and Tyr 284 on receptor activation by neurosteroids were quite specific because they did not markedly affect activation by piperidine-4-sulphonic acid, a partial agonist of the GABA_A receptor, or by the allosteric activator pentobarbitone

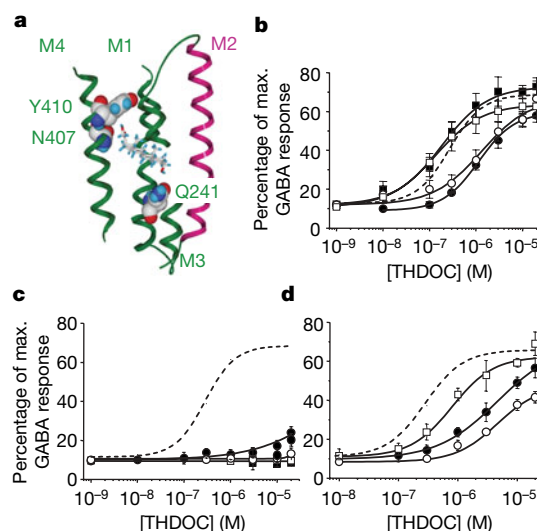


Figure 2 | Neurosteroid potentiation requires α -subunit M1 and M4 residues. **a**, Ribbon structure of α subunit M1–M4, showing α Gln 241, α Asn 407 and α Tyr 410 docking with a THDOC molecule, viewed from the lipid bilayer. The channel lining the M2 domain is shown in purple, and a section of M3 is omitted for clarity. **b**, **c**, THDOC concentration–response relationships for potentiation of EC₁₀ GABA responses on wild-type (α 1 β 2 γ 2, dashed line) and α 1 mutant receptors where Gln 241 is replaced with polar (**b**; filled squares, His; open squares, Asn; filled circles, Ser; open circles, Thr) or hydrophobic (**c**; filled squares, Ile; open squares, Leu; filled circles, Met; open circles, Trp) residues. **d**, Effects of mutating α Asn 407 and α Tyr 410 on THDOC potentiation. Dashed line, wild type; open circles, N407A Y410F; filled circles, Y410F; open squares, N407A. Error bars represent s.e.m.

(Supplementary Fig. 2a, c; Supplementary Table 2). This indicated that these residues might form a second neurosteroid-binding site, distinct from that mediating potentiation.

Although our data supported the existence of two distinct neurosteroid-binding sites on the GABA receptor, the mutation α Q241W not only abolished potentiation but also disrupted neurosteroid-induced activation. The apparent agonist efficacies for neurosteroids were reduced by α Q241W without affecting their potencies (Figs 1e and 3d, Supplementary Table 1). This raises the possibility that binding to α Thr 236– β Tyr 284 elicits only low-efficacy activation and that the high-efficacy activation, characteristic of wild-type receptors, depends on additional neurosteroid binding to α Gln 241– α Asn 407, thereby potentiating neurosteroid-induced activation. Disruption of the potentiation site should therefore attenuate, but not abolish, the efficacy of neurosteroid activation. Indeed, substitutions that significantly reduced the potentiation of GABA responses (for example α Q241S, α N407A and α Q241M) caused proportionate reductions in the apparent agonist efficacy of neurosteroids, whereas larger activation efficacies were observed for mutants (Q241H and Q241N) that supported greater levels of potentiation (Fig. 3c, d, and Supplementary Table 1). Overall, high-efficacy activation by neurosteroids seems to require occupancy of both the activation (α Thr 236, β Tyr 284) and potentiation (α Gln 241, α Asn 407) sites. Further evidence for two sites was provided by using the stereoisomer of ALLOP, 5β -pregnan- 3α -ol-20-one (5β 3 α), which is as efficacious as ALLOP and THDOC as a potentiator but is considerably less efficacious as an agonist (Supplementary Fig. 3). Because mutating α Gln 241– α Asn 407 and α Thr 236– β Tyr 284 reduced potentiation and activation by 5β 3 α , it must bind to the same residues as ALLOP and THDOC. However, whereas the potentiating site cannot distinguish between 5α 3 α and 5β 3 α isomers, the reduced agonist efficacy for

5β 3 α indicates that the activating site might have a different pharmacophore.

We then used our homology model to locate residues that line each site whose substitution might disrupt binding. We predicted that when bound to the potentiation site (α Gln 241 and α Asn 407), the hydrophobic A-ring of the neurosteroid is juxtaposed to the hydrophobic side chain of α Ile 238 (Fig. 4a). Introducing a similar-sized but polar side chain at residue 238 should cause electrostatic repulsion, reducing neurosteroid potency. As predicted, mutant I238N caused a rightward shift in the THDOC concentration–response curve for potentiation (Fig. 4c) without affecting the potentiation by diazepam (not shown). The apparent efficacy of neurosteroid activation was also reduced, which is consistent with high-efficacy neurosteroid activation requiring occupancy of both the activation and potentiation sites (Fig. 4d). Mutation of other residues lining this cavity, but not expected to disrupt neurosteroid binding, did not diminish potentiation (for example L231E and M235K; data not shown).

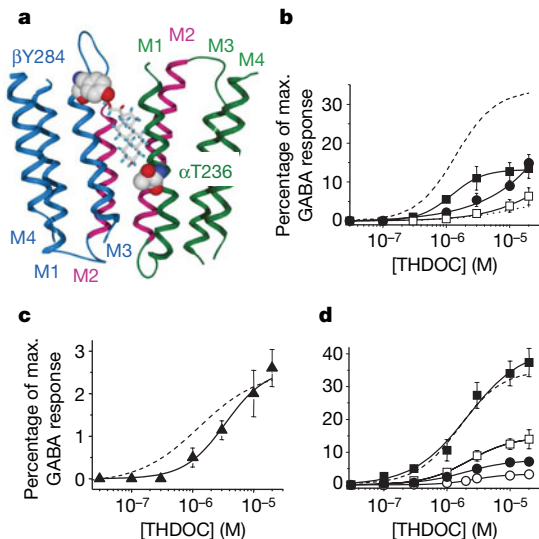


Figure 3 | Neurosteroid activation binding site spans the β / α -subunit interface. **a**, View through the lipid bilayer of model GABA_A receptor transmembrane region (extracellular and cytoplasmic domains removed) with a bound THDOC molecule. α Thr 236 and β Tyr 284 are predicted to face the surrounding lipid. **b**, THDOC concentration–response relationships for α 1 β 2 γ 2 receptor activation in the absence of GABA for the α T236 and β Y284 mutants. The wild-type (dashed) and α T236I (dotted) curves are taken from Fig. 1d. Filled squares, α T236S; open squares, α T236V; filled circles, β Y284F. **c**, Hydrophobic mutations of α Gln 241 reduce the apparent agonist efficacy of THDOC. Note that the concentration giving half-maximal response is unchanged from the wild type (dashed line). Dashed line, Q241W; triangles, Q241M. **d**, Mutations of α Gln 241 and α Asn 407 support varying degrees of direct receptor activation of α 1 β 2 γ 2 receptors by THDOC. Dashed line, wild type; filled squares, Q241H; open squares, Q241N; filled circles, N407A; open circles, Q241S. Error bars represent s.e.m.

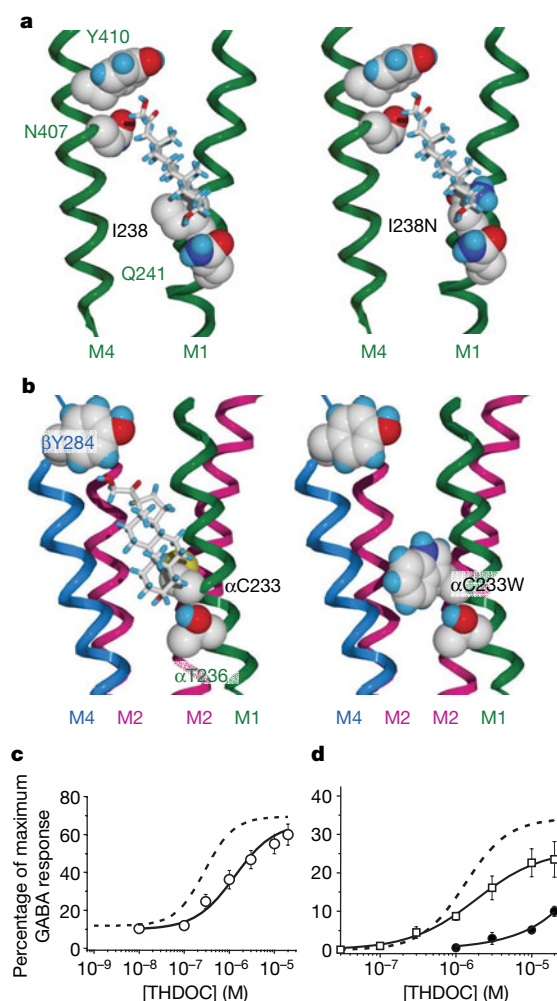


Figure 4 | Steric and dipolar disruption to the neurosteroid-binding sites. **a**, Homology model of THDOC bound to the potentiation site between M1 and M4 of the α subunit (M3 domain removed for clarity). Ile 238 is predicted to lie close to the A-ring of THDOC (left) and replacement with asparagine should repel the steroid (right). **b**, Cys 233 (left) is predicted to lie on the surface of the receptor; its replacement with tryptophan (right) increases the steric hindrance for THDOC binding to β Tyr 284 and α Thr 236. **c**, Effect of α I238N (circles) on the potency of THDOC potentiation of EC₁₀ GABA responses. Dashed line, wild type. **d**, Effect of C233W (filled circles) on the agonist potency of THDOC. Disruption of the potentiation site (α I238N; open squares) reduces THDOC apparent direct agonist efficacy. Dashed line, wild type. Error bars represent s.e.m.

The model also predicted that replacing C233 with tryptophan should sterically hinder neurosteroid binding to the activation site delineated by α Thr 236 and β Tyr 284 (Fig. 4b). Indeed, C233W substantially reduced the potency of neurosteroid activation (Fig. 4d) without affecting either the potentiation of GABA responses or receptor activation by GABA, piperidine-4-sulphonic acid or pentobarbitone (Supplementary Fig. 2a–c). This further indicates that α Thr 236– β Tyr 284 might form a second binding site from which neurosteroids initiate receptor activation. Substitution of vicinal residues (for example, α Pro 232), predicted not to line the putative binding site, had no effect on neurosteroid activation (data not shown). The mutations α L238N and α C233W therefore have specific effects on neurosteroid actions, each consistent with the disruption of distinct binding sites.

Our study shows that the activation and potentiation of GABA_A receptors by neurosteroids are mediated by two discrete groups of residues in the GABA_A receptor transmembrane domains: α Thr 236 and β Tyr 284 initiate activation, whereas α Gln 241 and α Asn 407 mediate the potentiation of responses to GABA or neurosteroids. It is also evident that significant receptor activation by neurosteroids relies on occupancy of both sites (Supplementary Fig. 3). The requirement for suitably spaced hydrogen-bonding residues at each site, and their specific disruption by mutating residues predicted to line the binding pockets, indicates that these residues might participate directly in neurosteroid binding. Given the receptor stoichiometry of $2\alpha:2\beta:1\gamma$, there will be two copies of these sites per receptor molecule (Supplementary Fig. 3). Moreover, these residues are conserved throughout the α and β subunit families and thus are ideally placed to mediate neurosteroid regulation at all GABA_A receptor subtypes. The identification of neurosteroid-binding sites on the GABA_A receptor will now enable the pathophysiological roles of these potent endogenous modulators of inhibitory neurotransmission to be clearly defined.

METHODS

Further details on the methods used in this study can be found in Supplementary Information.

Complementary DNA constructs and transfection. Murine α 1, β 2 and γ 2S subunits, chimaeras with RDL subunits, and all point mutants, were cloned into the plasmid pRK5. Functional GABA receptor expression was achieved by calcium phosphate transfection in human embryonic kidney cells.

Patch-clamp electrophysiology and data analysis. Whole-cell currents were recorded from single human embryonic kidney cells voltage-clamped at -40 mV, with an Axopatch 200B amplifier. Membrane currents were normalized to the maximal GABA response amplitude in each cell. Concentration–response curves were generated with the Hill equation. For studies of potentiation, neurosteroids and GABA were applied together without preincubation.

Homology modelling. The GABA_A receptor α 1 and β 2 subunit transmembrane domains were modelled on a cryo-electron microscopic image of the homologous nicotinic acetylcholine receptor M1–M4 domains (Protein Data Bank accession number 1OED)²³ with the use of Deep View²⁷. The GABA_A receptor transmembrane domains were aligned such that residues believed to contribute to volatile-anaesthetic-binding sites faced into the cavity between the four transmembrane domains where these compounds are thought to bind²⁸. Residue Gln 241 lies on the same side of the M1 helix as Leu 231, a known member of this site. The models were energy-minimized with the GROMOS 43B1 force field and the positions of side chains after mutations were assessed for conformational integrity with the use of Ramachandran plot analysis. For illustrative purposes, ALLOP and THDOC molecules were docked by eye and are not intended to predict the molecules' absolute orientations within the proposed binding cavities.

Received 16 May; accepted 9 October 2006.

Published online 15 November 2006.

1. Belelli, D. & Lambert, J. J. Neurosteroids: endogenous regulators of the GABA_A receptor. *Nature Rev. Neurosci.* **6**, 565–575 (2005).
2. Reddy, D. S. Is there a physiological role for the neurosteroid THDOC in stress-sensitive conditions? *Trends Pharmacol. Sci.* **24**, 103–106 (2003).

3. Stoffel-Wagner, B. Neurosteroid biosynthesis in the human brain and its clinical implications. *Ann. NY Acad. Sci.* **1007**, 64–78 (2003).
4. Kumar, S., Fleming, R. L. & Morrow, A. L. Ethanol regulation of gamma-aminobutyric acid A receptors: genomic and nongenomic mechanisms. *Pharmacol. Ther.* **101**, 211–226 (2004).
5. Backstrom, T. et al. Pathogenesis in menstrual cycle-linked CNS disorders. *Ann. NY Acad. Sci.* **1007**, 42–53 (2003).
6. Finn, D. A., Ford, M. M., Wren, K. M., Roselli, C. E. & Crabbe, J. C. The role of pregnane neurosteroids in ethanol withdrawal: behavioral genetic approaches. *Pharmacol. Ther.* **101**, 91–112 (2004).
7. Eser, D. et al. Neuroactive steroids as modulators of depression and anxiety. *Neuroscience* **138**, 1041–1048 (2006).
8. Marx, C. E. et al. Neuroactive steroids are altered in schizophrenia and bipolar disorder: relevance to pathophysiology and therapeutics. *Neuropsychopharmacology* **31**, 1249–1263 (2006).
9. Fritschy, J. M. & Brunig, I. Formation and plasticity of GABAergic synapses: physiological mechanisms and pathophysiological implications. *Pharmacol. Ther.* **98**, 299–323 (2003).
10. Purdy, R. H., Morrow, A. L., Moore, P. H. Jr & Paul, S. M. Stress-induced elevations of γ -aminobutyric acid type A receptor-active steroids in the rat brain. *Proc. Natl Acad. Sci. USA* **88**, 4553–4557 (1991).
11. Barbaccia, M. L. et al. The effects of inhibitors of GABAergic transmission and stress on brain and plasma allopregnanolone concentrations. *Br. J. Pharmacol.* **120**, 1582–1588 (1997).
12. Zhu, W. J. & Vicini, S. Neurosteroid prolongs GABA_A channel deactivation by altering kinetics of desensitized states. *J. Neurosci.* **17**, 4022–4031 (1997).
13. Stell, B. M., Brickley, S. G., Tang, C. Y., Farrant, M. & Mody, I. Neuroactive steroids reduce neuronal excitability by selectively enhancing tonic inhibition mediated by δ subunit-containing GABA_A receptors. *Proc. Natl Acad. Sci. USA* **100**, 14439–14444 (2003).
14. Belelli, D. & Herd, M. B. The contraceptive agent Provera enhances GABA_A receptor-mediated inhibitory neurotransmission in the rat hippocampus: evidence for endogenous neurosteroids? *J. Neurosci.* **23**, 10013–10020 (2003).
15. Majewska, M. D., Harrison, N. L., Schwartz, R. D., Barker, J. L. & Paul, S. M. Steroid hormone metabolites are barbiturate-like modulators of the GABA receptor. *Science* **232**, 1004–1007 (1986).
16. Rick, C. E., Ye, Q., Finn, S. E. & Harrison, N. L. Neurosteroids act on the GABA_A receptor at sites on the N-terminal side of the middle of TM2. *Neuroreport* **9**, 379–383 (1998).
17. Akk, G. et al. Neurosteroid access to the GABA_A receptor. *J. Neurosci.* **25**, 11605–11613 (2005).
18. Ueno, S., Tsutsui, M., Toyohira, Y., Minami, K. & Yanagihara, N. Sites of positive allosteric modulation by neurosteroids on ionotropic gamma-aminobutyric acid receptor subunits. *FEBS Lett.* **566**, 213–217 (2004).
19. Chen, R. et al. Cloning and functional expression of a *Drosophila* γ -aminobutyric acid receptor. *Proc. Natl Acad. Sci. USA* **91**, 6069–6073 (1994).
20. Harrison, N. L., Majewska, M. D., Harrington, J. W. & Barker, J. L. Structure-activity relationships for steroid interaction with the gamma-aminobutyric acid-A receptor complex. *J. Pharmacol. Exp. Ther.* **241**, 346–353 (1987).
21. Grishkovskaya, I. et al. Crystal structure of human sex hormone-binding globulin: steroid transport by a laminin G-like domain. *EMBO J.* **19**, 504–512 (2000).
22. Brzozowski, A. M. et al. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–758 (1997).
23. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
24. Williams, D. B. & Akabas, M. H. γ -Aminobutyric acid increases the water accessibility of M3 membrane-spanning segment residues in γ -aminobutyric acid type A receptors. *Biophys. J.* **77**, 2563–2574 (1999).
25. Lobo, I. A., Mascia, M. P., Trudell, J. R. & Harris, R. A. Channel gating of the glycine receptor changes accessibility to residues implicated in receptor potentiation by alcohols and anesthetics. *J. Biol. Chem.* **279**, 33919–33927 (2004).
26. Jung, S., Akabas, M. H. & Harris, R. A. Functional and structural analysis of the GABA_A receptor α 1 subunit during channel gating and alcohol modulation. *J. Biol. Chem.* **280**, 308–316 (2005).
27. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
28. Trudell, J. R. & Bertaccini, E. Comparative modeling of a GABA_A α 1 receptor using three crystal structures as templates. *J. Mol. Graph. Model.* **23**, 39–49 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Duguid, P. Miller and P. Thomas for comments on the manuscript. This work was supported by the Medical Research Council and The Wellcome Trust.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.G.S. (t.smart@ucl.ac.uk) and A.M.H. (alastair.hosie@ucl.ac.uk).

LETTERS

Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen

Barend H. J. de Graaf^{1*}, Jason J. Rudd^{1*†}, Michael J. Wheeler^{1*†}, Ruth M. Perry¹, Elizabeth M. Bell¹, Kim Osman¹, F. Christopher H. Franklin¹ & Veronica E. Franklin-Tong¹

In higher plants, sexual reproduction involves interactions between pollen and pistil. A key mechanism to prevent inbreeding is self-incompatibility through rejection of incompatible ('self') pollen¹. In *Papaver rhoeas*, S proteins encoded by the stigma interact with incompatible pollen, triggering a Ca^{2+} -dependent signalling network^{2–5} resulting in pollen tube inhibition and programmed cell death⁶. The cytosolic phosphoprotein p26.1, which has been identified in incompatible pollen, shows rapid, self-incompatibility-induced Ca^{2+} -dependent hyperphosphorylation *in vivo*³. Here we show that p26.1 comprises two proteins, Pr-p26.1a and Pr-p26.1b, which are soluble inorganic pyrophosphatases (sPPases). These proteins have classic Mg^{2+} -dependent sPPase activity, which is inhibited by Ca^{2+} , and unexpectedly can be phosphorylated *in vitro*. We show that phosphorylation inhibits sPPase activity, establishing a previously unknown mechanism for regulating eukaryotic sPPases. Reduced sPPase activity is predicted to result in the inhibition of many biosynthetic pathways, suggesting that there may be additional mechanisms of self-incompatibility-mediated pollen tube inhibition. We provide evidence that sPPases are required for growth and that self-incompatibility results in an increase in inorganic pyrophosphate, implying a functional role for Pr-p26.1.

Inorganic pyrophosphatases (PPases) are highly conserved across prokaryotes and eukaryotes. They are enzymes that, through hydrolysis of inorganic pyrophosphate (PPi), provide the driving force for many metabolic reactions. PPi is generated during biopolymer (for example, cellulose) synthesis and hydrolysed to inorganic phosphate (2Pi); this reaction provides a thermodynamic pull favouring biosynthesis^{7,8}.

The self-incompatibility (SI) response in pollen of *Papaver* provides a model system in which to investigate signalling in plant cells. Several targets of SI-induced Ca^{2+} signals have been identified, including the F-actin cytoskeleton^{4,5} and a programmed cell death signalling cascade⁶. A rapid increase in SI-specific phosphorylation of a cytosolic pollen protein, p26.1, in incompatible pollen undergoing SI *in vivo*, identified this protein as being involved in SI (ref. 3).

To determine the nature of p26.1, we cloned the complementary DNA encoding it from pollen, obtaining two full-length sequences, Pr-p26.1a and Pr-p26.1b (Supplementary Fig. S2). Pr-p26.1a is predicted to encode a 24.4-kDa protein, pI 6.11, comprising 217 amino acids; Pr-p26.1b is predicted to encode a 26.5-kDa protein, pI 6.03, comprising 236 amino acids. Both predicted amino acid sequences have high homology to Family I sPPases. The protein encoded by Pr-p26.1a contains peptide sequence obtained from p26.1 (Supplementary Fig. S2). At the amino acid level, Pr-p26.1a and Pr-p26.1b have 78% identity. There is considerable evolutionary

conservation of Family I sPPases^{8,9}; bacterial sPPases and Pr-p26.1a or Pr-p26.1b have 57% overall identity. Both Pr-p26.1a and Pr-p26.1b possess the classic sPPase signature sequence DXDXXDX with three aspartate residues comprising a divalent-ion-binding site (usually for Mg^{2+}) essential for catalytic activity⁸, 14 conserved residues common to all sPPases, and putative phosphorylation sites (Supplementary Fig. S2).

Although their predicted molecular masses differ, the endogenous Pr-p26.1a and Pr-p26.1b proteins co-migrated on SDS polyacrylamide gel electrophoresis (Fig. 1a). We found that expression is limited primarily to mature pollen (Fig. 1a). The lack of crossreactivity in cytosolic extracts from other tissues suggested that they contain low concentrations of sPPases. These data indicate a functional involvement for Pr-p26.1a and Pr-p26.1b in pollen-specific processes, during either pollen germination or pollen tube growth, because there is little evidence for their expression during pollen development.

Although several cytosolic sPPases have been cloned and characterized from plants^{10–12}, examples of these are rare; indeed, it has been stated that the plant cytosol lacks sPPases¹³. Plant sPPases are

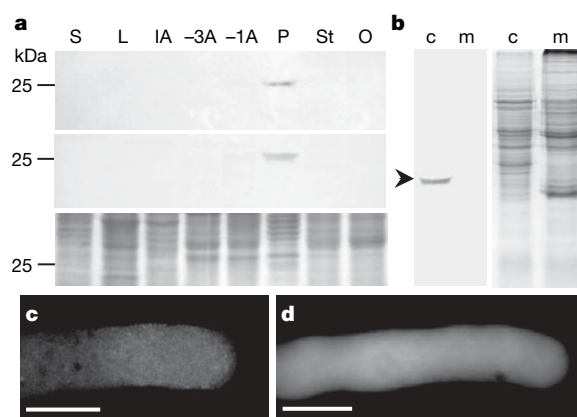


Figure 1 | Expression analysis of Pr-p26.1a and Pr-p26.1b sPPases.

a, Western blots of cytosolic proteins (30 µg) from *Papaver* stem (S), leaf (L), immature anthers (IA), stage -3 (-3A) and stage -1 (-1A) anthers, hydrated pollen (P), stigma (St) and ovary (O). Antibody to p26.1a crossreacts specifically with Pr-p26.1a in pollen (top). Antibody to p26.1 crossreacts with a doublet in pollen (middle). Coomassie blue staining shows equal loading (bottom). **b**, Fractionated pollen proteins, probed with antibody to p26.1, show Pr-p26.1 proteins in cytosolic (c) but not microsomal (m) fractions (left). Coomassie blue staining shows equal loading (right). **c**, **d**, Immunolocalization with antibody to p26.1 (**c**) and Pr-p26.1b-GFP expression (**d**) show that Pr-p26.1a and Pr-p26.1b proteins in pollen tubes are cytosolic. Scale bar, 10 µm.

¹School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [†]Present addresses: Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK (J.J.R.); Institute of Cancer Studies, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK (M.J.W.).

*These authors contributed equally to this work.

localized primarily to plastids rather than the cytosol¹⁴. As the SI-stimulated phosphoprotein p26.1 was originally identified in cytosolic extracts³, we fractionated pollen protein extracts to confirm its localization. Pr-p26.1a and Pr-p26.1b were detected only in the soluble fraction (Fig. 1b). Immunolocalization and green fluorescent protein (GFP) fusion proteins confirmed that these sPPases were cytosolic (Fig. 1c, d). A possible explanation for the unusual cytosolic localization of these sPPases is that they are abundant in metaboli-

cally highly active cells. Notably, an sPPase has been purified from the latex of *Hevea brasiliensis*¹⁵ that is involved in rubber biosynthesis. In addition, overexpression of a bacterial sPPase in the cytosol of potato tubers increases starch synthesis¹⁶. Thus, pollen tubes may express cytosolic sPPases because they require high metabolic activity to generate new membrane and cell wall for pollen tube extension.

To verify that the proteins encoded by *Pr-p26.1a* and *Pr-p26.1b* are sPPases, we tested whether the recombinant proteins had the key properties of sPPases by measuring their sPPase specific activities⁸. Both recombinant Pr-p26.1a and Pr-p26.1b had high Mg^{2+} -dependent sPPase specific activities of $2,320 \pm 140$ and $2,170 \pm 160 \mu\text{mol Pi mg}^{-1} \text{min}^{-1}$ (mean $\pm$ s.e.m., $n = 4$), respectively (Fig. 2a). Almost no sPPase activity was detected when Mg^{2+} was chelated by EDTA. Other divalent cations (Ca^{2+} , Mn^{2+}) could not substitute for Mg^{2+} . Ca^{2+} inhibited sPPase activity, reducing the specific activity of Pr-p26.1a and Pr-p26.1b by $53.2 \pm 1.52\%$ and $50.4 \pm 2.50\%$, respectively ($n = 4$, $P < 0.001$). Mn^{2+} could not substitute for the effect of Ca^{2+} . This finding unequivocally shows that Pr-p26.1a and Pr-p26.1b have classic Mg^{2+} -dependent sPPase activity that is inhibited by Ca^{2+} . The two proteins show indistinguishable sPPase specific activities ($P = 0.617$, NS).

We tested the hypothesis that cytosolic sPPases might be abundant only in cells with high metabolic activity by comparing sPPase activities in cytosolic extracts from *Papaver* pollen with those from other tissues. The mean sPPase specific activity in germinated pollen was $36.8 \pm 0.83 \mu\text{mol Pi mg}^{-1} \text{min}^{-1}$ ($n = 4$), 7.8-fold higher than that in mature leaf (sPPase specific activity of $4.7 \pm 0.15 \mu\text{mol Pi mg}^{-1} \text{min}^{-1}$, $n = 4$, $P < 0.001$). Other tissues showed low activity (Supplementary Fig. S3). This finding supports our hypothesis that pollen is unusual in having high cytosolic sPPase activity.

As p26.1 was identified through its SI-induced phosphorylation, we examined whether Pr-p26.1a and Pr-p26.1b isolated from pollen cytosol are phosphorylated by using a phosphocolumn to select phosphorylated proteins. Western blotting showed that the phosphorylated fraction contained Pr-p26.1a and Pr-p26.1b (Fig. 2b). Matrix-assisted laser desorption-ionization quadrupole time-of-flight (MALDI-QTOF) tandem mass spectrometry identified three peptides that matched Pr-p26.1a and Pr-p26.1b sequences (Supplementary Fig. S2), indicating that both sPPases are present in pollen and phosphorylated. To establish unequivocally that Pr-p26.1a and Pr-p26.1b can be phosphorylated, we carried out *in vitro* kinase assays using recombinant His-tagged sPPase proteins. His-affinity purification showed that both Pr-p26.1a and Pr-p26.1b are phosphorylated (Fig. 2c) in a Ca^{2+} -dependent manner (Fig. 2d). Taken together with our previous identification of p26.1 (ref. 3), this observation seems to represent, to our knowledge, the first example of the phosphorylation of eukaryotic sPPases involved in a physiological process. Although examples of kinase-dependent phosphorylation of sPPases in prokaryotes exist^{17,18}, the only demonstration of phosphorylation of a eukaryotic sPPase has been *in vitro*¹⁹; thus, the biological significance of this modification is unknown.

Protein phosphorylation is an important regulatory mechanism for altering enzyme activity. Because SI induces Ca^{2+} -dependent phosphorylation of p26.1 (ref. 3), we ascertained whether phosphorylation could modify sPPase activity. The sPPase specific activities of phosphorylated endogenous pollen proteins containing phosphorylated Pr-p26.1 were compared with those of unphosphorylated extracts containing equivalent amounts of Pr-p26.1 (Fig. 2e). Phosphorylated Pr-p26.1 showed a reduction in sPPase specific activity of 59.3% as compared with equivalent unphosphorylated extracts. Thus, phosphorylation of Pr-p26.1 markedly reduces its sPPase activity. As expected, both phosphorylated and non-phosphorylated forms of Pr-p26.1 showed Ca^{2+} -dependent inhibition of sPPase activity (Fig. 2e), with a reduction in specific activity of $62.0 \pm 1.5\%$ and $59 \pm 3.0\%$, respectively. Statistical analysis showed that both phosphorylation and Ca^{2+} had highly significant effects

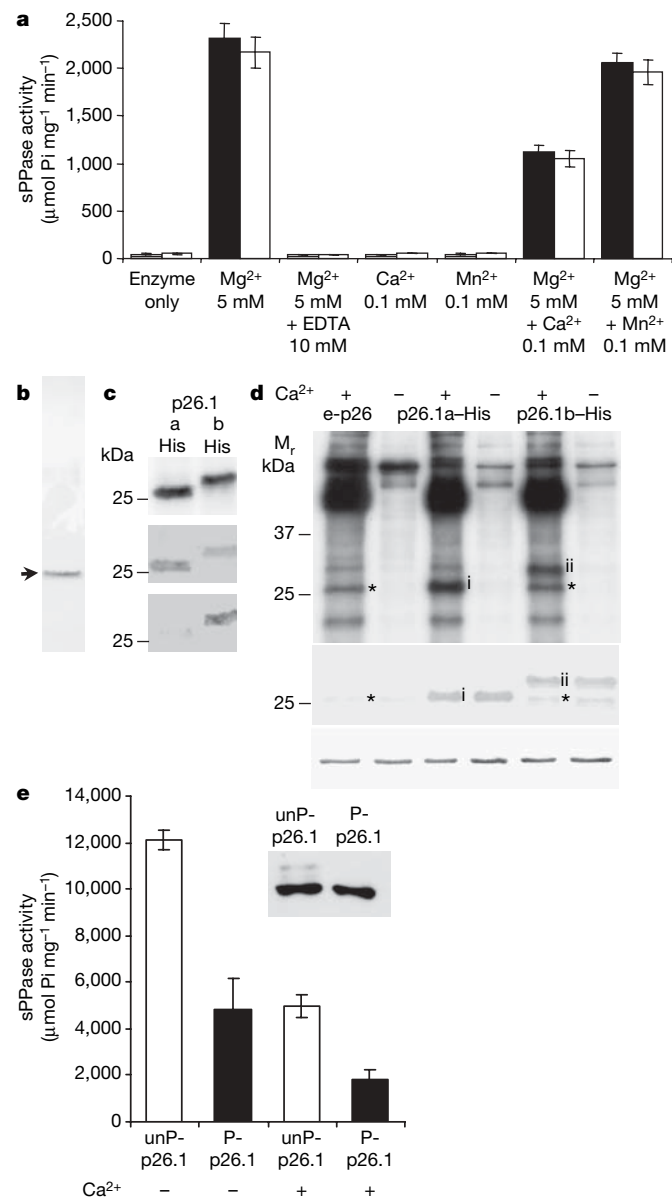


Figure 2 | Pr-p26.1a and Pr-p26.1b show sPPase activity and are phosphorylated, and phosphorylation affects sPPase activity. **a**, sPPase activities of Pr-p26.1a-His (filled bars) and Pr-p26.1b-His (open bars). Data are the mean $\pm$ s.e.m. ($n = 4$). **b**, Phosphorylated pollen proteins crossreact with antibody to p26.1 (arrow). **c**, Autoradiography shows that Pr-p26.1a-His and Pr-p26.1b-His proteins are phosphorylated (top). Coomassie blue staining identifies Pr-p26.1a and Pr-p26.1b (middle). Western blot probed with His-tag antisera identifies these proteins as p26.1a-His and p26.1b-His (bottom). **d**, Autoradiograph (top) and Western blot probed with antibody to p26.1 (middle) show Ca^{2+} -dependent *in vitro* phosphorylation of endogenous p26.1 (asterisk), Pr-p26.1a-His (i) and Pr-p26.1b-His (ii) in the presence or absence of Ca^{2+} . Anti-actin staining shows equal loading (bottom). **e**, Phosphorylated and unphosphorylated Pr-p26.1 protein (inset) were assayed for sPPase activity in the presence and absence of Ca^{2+} . Open bars, unphosphorylated p26.1; filled bars, phosphorylated p26.1. Data are the mean $\pm$ s.e.m. ($n = 3$).

($n = 3$ for both, $P < 0.001$) on sPPase activity; the effects of phosphorylation and Ca^{2+} were not independent ($P = 0.027$). Our data thus show that Pr-p26.1 sPPase activity is markedly reduced by phosphorylation and that Ca^{2+} also reduces its activity.

This demonstration that phosphorylation inhibits sPPase activity establishes a previously unknown mechanism for regulating sPPase catalytic activity in eukaryotes. To our knowledge, signalling to sPPases in a eukaryote has not previously been demonstrated. So far, the only indication of a role for sPPase phosphorylation has been in regulation of purine biosynthesis in bacteria¹⁸; however, the effect of phosphorylation on sPPase activity was not ascertained in that study. Thus, our data provide insight into the regulation of sPPases, implicating a key role for kinase-dependent phosphorylation.

These data suggest a mechanism for regulating incompatible pollen tube growth. The ubiquitous function of sPPases is axiomatic. sPPases are important enzymes in biosynthesis⁷, and extensive evidence indicates that inhibition of sPPases decreases metabolic activity. sPPase activity is essential for growth in bacteria²⁰ and yeast²¹, and increasing sPPase activity in the plant cytosol (where it is not usually located) results in an increase in starch²². Pollen tube growth requires extensive biosynthesis of membrane and cell wall components to enable it to grow. In SI, inhibition of biosynthetic activity could clearly contribute to the arrest of growth required to prevent self-fertilization. To confirm the functional requirement of Pr-p26.1 sPPases in pollen tube growth, we used an antisense oligonucleotide approach^{23,24} to downregulate Pr-p26.1. We observed a significant reduction in pollen tube length ($79.8 \pm 1.1\%$) in the presence of antisense oligonucleotides targeted to the Pr-p26.1a and Pr-p26.1b sequence ($n = 3$, $P < 0.001$), but not in the presence of the corresponding sense oligonucleotides ($n = 3$, $P = 0.058$, NS; Fig. 3a and Supplementary Fig. S4). This shows that p26.1 is crucial for pollen tube growth.

To provide a functional link between changes in sPPase activity and SI induction in pollen, we measured the concentration of PPI after SI induction. We considered that if sPPase activity is reduced by SI induction, then the cellular PPI concentration would increase, because sPPases remove excess PPI by hydrolysing it to 2Pi. Normally growing pollen tubes had a PPI concentration of $38.2 \pm 5.2 \text{ pmol mg}^{-1} \text{ pollen}^{-1}$ ($n = 6$), whereas the PPI concentration in SI-induced incompatible pollen tubes was 180% of the control value ($n = 9$; Fig. 3b). The increase in PPI in SI-induced pollen as compared with untreated controls ($P = 0.001$) is consistent with the prediction that SI results in high PPI owing to a decrease in sPPase activity. This observation provides evidence for a mechanistic link between sPPases and SI *in vivo* in pollen tubes.

In summary, we have demonstrated phosphorylation of the sPPases Pr-p26.1a and Pr-p26.1b and provide evidence that this phosphorylation inhibits sPPase activity. Our data identify a regulatory mechanism for the inhibition of this important class of enzyme.

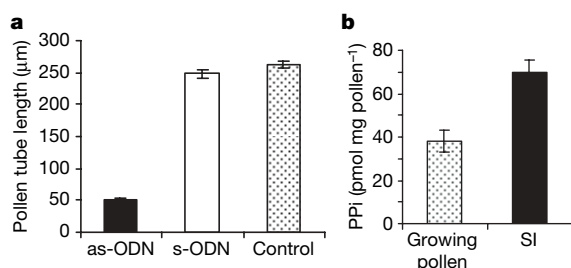


Figure 3 | Functional evidence for Pr-p26.1 and sPPase involvement in modulating pollen tube growth. **a**, Effect of Pr-p26.1 as-ODN on pollen tube growth. Antisense oligodeoxynucleotide inhibited growth, whereas s-ODN or cytofectin alone (control) did not. Twenty-five pollen tubes were measured in each of three independent experiments. **b**, PPI assay of normally growing pollen tubes ($n = 6$) and SI-induced pollen tubes ($n = 9$). Data are the mean $\pm$ s.e.m.

Furthermore, we have provided evidence showing its importance in regulating pollen tube growth. Because pollen tubes require high metabolic activity for extensive biosynthesis, which will release PPI, the inhibition of sPPase activity is expected to be especially effective in contributing to growth arrest (Supplementary Fig. S1), making it an ideal mechanism for operating in SI, which depends on pollen tube inhibition to prevent self-fertilization. Thus, identification of sPPases as a target for the SI signals in *Papaver* pollen represents a significant advance in our understanding of SI. The earliest SI response is an increase in intracellular Ca^{2+} (refs 2, 25), which triggers a signalling network to stop pollen tube growth. Actin depolymerization⁵ will rapidly inhibit pollen tube growth, and programmed cell death^{6,26} will ensure that pollen tube growth does not resume. As we have shown here, sPPase modification also has a key functional role in SI by regulating pollen tube growth.

METHODS

Cytosolic protein extraction. Cytosolic proteins were extracted from pollen and other tissues as described³.

In vitro kinase reactions. Pollen cytosolic extracts³ were phosphorylated *in vitro* by standard procedures³; phosphorylation of the recombinant Pr-p26.1a and Pr-p26.1b His-tagged proteins was carried out in the presence of cytosolic pollen protein extracts (see Supplementary Methods for full details).

sPPase assays. To determine recombinant sPPase activity, recombinant Pr-p26.1a and Pr-p26.1b aliquots were assayed for sPPase activity as described¹¹ and are expressed as $\mu\text{mol Pi per mg Pr-p26.1a/Pr-p26.1b per min}$ ($\mu\text{mol Pi mg}^{-1} \text{ min}^{-1}$). For endogenous pollen and tissue sPPase activities, cytosolic extracts³ were assayed¹¹ and are expressed as $\mu\text{mol Pi per mg total protein per min}$. Details of determination of phosphorylated endogenous pollen sPPase activities are described in Supplementary Methods. Some pollen cytosolic extracts³ were phosphorylated *in vitro* and subjected to phosphocolumn purification (see Supplementary Methods).

Antisense oligonucleotide experiments. A phosphorothioated antisense oligodeoxynucleotide (as-ODN) and its sense control (s-ODN) were designed to downregulate both genes (see Supplementary Methods for sequences). Pollen was grown *in vitro* as described⁶ and treated with as-ODN and s-ODN as described²⁴ (see Supplementary Methods for full details).

PPI assays. Pollen was grown *in vitro* and SI induced as described⁶. PPI assays were carried out with a PPIper pyrophosphate assay kit (Molecular Probes) in accordance with the manufacturer's instructions (see Supplementary Methods for full details).

Statistics. Statistical tests were carried out with MINITAB. Tests on most data were two-way comparisons between pairs of data using Student's *t*-test. Two-way analyses of variance were carried out on sPPase activity data from Fig. 2a, e.

Additional materials and methods, including cloning, antisera, western blotting, immunolocalization, transient expression using particle bombardment, imaging, His-tag affinity purification of recombinant proteins, phosphocolumn affinity purification, mass spectrometry are described in Supplementary Methods.

Received 28 July; accepted 4 October 2006.

Published online 5 November 2006.

- McClure, B. A. & Franklin-Tong, V. E. Gametophytic self-incompatibility: understanding the cellular mechanisms involved in 'self' pollen tube inhibition. *Planta* **224**, 233–245 (2006).
- Franklin-Tong, V. E., Ride, J. P., Read, N. D., Trewavas, A. J. & Franklin, F. C. H. The self-incompatibility response in *Papaver rhoeas* is mediated by cytosolic-free calcium. *Plant J.* **4**, 163–177 (1993).
- Rudd, J. J., Franklin, F. C. H., Lord, J. M. & Franklin-Tong, V. E. Increased phosphorylation of a 26-kD pollen protein is induced by the self-incompatibility response in *Papaver rhoeas*. *Plant Cell* **8**, 713–724 (1996).
- Geitmann, A., Snowman, B. N., Emons, A. M. C. & Franklin-Tong, V. E. Alterations in the actin cytoskeleton of pollen tubes are induced by the self-incompatibility reaction in *Papaver rhoeas*. *Plant Cell* **12**, 1239–1251 (2000).
- Snowman, B. N., Kovar, D. R., Shevchenko, G., Franklin-Tong, V. E. & Staiger, C. J. Signal-mediated depolymerization of actin in pollen during the self-incompatibility response. *Plant Cell* **14**, 2613–2626 (2002).
- Thomas, S. G. & Franklin-Tong, V. E. Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature* **429**, 305–309 (2004).
- Kornberg, A. *On the Metabolic Significance of Phosphorylation and Pyrophosphorylation* (eds Kasha, H. & Pullman, P.) (Academic Press, New York, 1962).
- Cooperman, B. S., Baykov, A. A. & Lahti, R. Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Trends Biochem. Sci.* **17**, 262–266 (1992).

9. Sivula, T. *et al.* Evolutionary aspects of inorganic pyrophosphatase. *FEBS Lett.* **454**, 75–80 (1999).
10. du Jardin, P., Rojas-Beltrán, J., Gebhardt, C. & Brasseur, R. Molecular cloning and characterization of a soluble inorganic pyrophosphatase in potato. *Plant Physiol.* **109**, 853–860 (1995).
11. Visser, K., Heimovaara-Dijkstra, S., Kijne, J. W. & Wang, M. Molecular cloning and characterization of an inorganic pyrophosphatase from barley. *Plant Mol. Biol.* **37**, 131–140 (1998).
12. Rojas-Beltrán, J. A. *et al.* Identification of cytosolic Mg^{2+} -dependent soluble inorganic pyrophosphatases in potato and phylogenetic analysis. *Plant Mol. Biol.* **39**, 449–461 (1999).
13. Plaxton, W. C. *Plant Response to Stress: Biochemical Adaptations to Phosphate Deficiency* (Marcel Dekker, New York, 2004).
14. Gross, P. & ap Rees, T. Alkaline inorganic pyrophosphatase and starch synthesis in amyloplasts. *Planta* **167**, 140–145 (1986).
15. Jacob, J.-L., Prevot, J.-C., Clement-Vidal, A. & d'Auzac, J. Inorganic pyrophosphate metabolism in *Hevea brasiliensis* latex. Characteristics of cytosolic alkaline pyrophosphatase. *Plant Physiol. Biochem.* **27**, 355–364 (1989).
16. Geigenberger, P. *et al.* Overexpression of pyrophosphatase leads to increased sucrose degradation and starch synthesis, increased activities of enzymes for sucrose-starch interconversions, and increased levels of nucleotides in growing potato tubers. *Planta* **205**, 428–437 (1998).
17. Rajagopal, L., Clancy, A. & Rubens, C. E. A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J. Biol. Chem.* **278**, 14429–14441 (2003).
18. Rajagopal, L., Vo, A., Silvestroni, A. & Rubens, C. E. Regulation of purine biosynthesis by a eukaryotic-type kinase in *Streptococcus agalactiae*. *Mol. Microbiol.* **56**, 1329–1346 (2005).
19. Vener, A. V., Smirnova, I. N. & Baykov, A. A. Phosphorylation of rat liver inorganic pyrophosphatase by ATP in the absence and in the presence of protein kinase. *FEBS Lett.* **264**, 40–42 (1990).
20. Chen, J. *et al.* Pyrophosphatase is essential for growth of *Escherichia coli*. *J. Bacteriol.* **172**, 5686–5689 (1990).
21. Perez-Castineira, J. R., Lopez-Marques, R. L., Villalba, J. M., Losada, M. & Serrano, A. Functional complementation of yeast cytosolic pyrophosphatase by bacterial and plant H^{+} -translocating pyrophosphatases. *Proc. Natl Acad. Sci. USA* **99**, 15914–15919 (2002).
22. Sonnewald, U. Expression of *E. coli* inorganic pyrophosphatase in transgenic plants alters photoassimilate partitioning. *Plant J.* **2**, 571–581 (1992).
23. Estruch, J., Kadwell, S., Merlin, E. & Crossland, L. Cloning and characterization of a maize pollen-specific calcium-dependent calmodulin-independent protein kinase. *Proc. Natl Acad. Sci. USA* **91**, 8837–8841 (1994).
24. Moutinho, A. *et al.* Antisense perturbation of protein function in living pollen tubes. *Sex. Plant Reprod.* **14**, 101–104 (2001).
25. Franklin-Tong, V. E., Hackett, G. & Hepler, P. K. Ratio-imaging of $[Ca^{2+}]_i$ in the self-incompatibility response in pollen tubes of *Papaver rhoeas*. *Plant J.* **12**, 1375–1386 (1997).
26. Thomas, S. G., Huang, S., Li, S., Staiger, C. J. & Franklin-Tong, V. E. Actin depolymerization is sufficient to induce programmed cell death in self-incompatible pollen. *J. Cell Biol.* **174**, 221–229 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Wullems for the *ntp303* pollen promoter; G. Jones for help and advice on statistical analysis; and E. Sanchez-Moran for help with imaging. Work in the laboratories of F.C.H.F. and V.E.F.-T. is funded by the UK Biotechnology and Biological Sciences Research Council.

Author Information Sequences have been deposited in the EMBL Nucleotide Sequence Database (<http://www.ebi.ac.uk/embl/>) under accession codes AM162550 and AM162551. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.E.F.-T. (v.e.franklin-tong@bham.ac.uk).

LETTERS

Polo kinase controls cell-cycle-dependent transcription by targeting a coactivator protein

Zoulfia Darieva¹, Richard Bulmer², Aline Pic-Taylor^{2†}, Kathryn S. Doris², Marco Geymonat³, Steven G. Sedgwick³, Brian A. Morgan² & Andrew D. Sharrocks¹

Polo kinases have crucial conserved functions in controlling the eukaryotic cell cycle through orchestrating several events during mitosis^{1,2}. An essential element of cell cycle control is exerted by altering the expression of key regulators³. Here we show an important function for the polo kinase Cdc5p in controlling cell-cycle-dependent gene expression that is crucial for the execution of mitosis in the model eukaryote *Saccharomyces cerevisiae*. In particular, we find that Cdc5p is temporally recruited to promoters of the cell-cycle-regulated *CLB2* gene cluster, where it targets the Mcm1p–Fkh2p–Ndd1p transcription factor complex, through direct phosphorylation of the coactivator protein Ndd1p. This phosphorylation event is required for the normal temporal expression of cell-cycle-regulated genes such as *CLB2* and *SWI5* in G2/M phases. Furthermore, severe defects in cell division occur in the absence of Cdc5p-mediated phosphorylation of Ndd1p. Thus, polo kinase is required for the production of key mitotic regulators, in addition to previously defined roles in controlling other mitotic events.

In *S. cerevisiae* the polo kinase Cdc5p has been implicated in several processes associated with mitosis, including mitotic entry, spindle assembly, cytokinesis and mitotic exit². Although several direct targets for Cdc5p have been identified, including Swe1p, Scc1p and Bfa1p^{4–7}, no direct links have been made between polo kinases and the transcription factor complexes that control cell-cycle-dependent transcription.

In *S. cerevisiae* transcriptional regulation of the *CLB2* gene cluster in the G2 and M phases represents a key cell cycle control point. This gene cluster encodes several factors required for mitotic progression including the B-type cyclin Clb2p, transcriptional regulators such as Swi5p, and Cdc5p polo kinase^{3,8}. A major determinant of this co-ordinated gene expression is a transcription factor complex that consists of Mcm1p, the forkhead transcription factor Fkh2p and the coactivator protein Ndd1p^{3,9–13}. The regulation of the Mcm1p–Fkh2p–Ndd1p complex is linked to the cell cycle through the sequential phosphorylation of Fkh2p by the Clb5p–Cdc28p¹⁴ kinase complex, and Ndd1p by the Clb2p–Cdc28p^{15,16} kinase complex. Although these regulatory phosphorylation events promote Ndd1p recruitment and target gene activation, it is less clear whether they alone account for the complete regulation of the Mcm1p–Fkh2p–Ndd1p complex. We therefore investigated whether the polo kinase Cdc5p has a role in controlling the activity of the Mcm1p–Fkh2p–Ndd1p complex.

First we investigated whether we could detect Cdc5p at the promoters of genes within the *CLB2* cluster by chromatin immunoprecipitation (ChIP) analysis. We used a strain overexpressing Cdc5p to permit the analysis of Cdc5p recruitment at times when endogenous

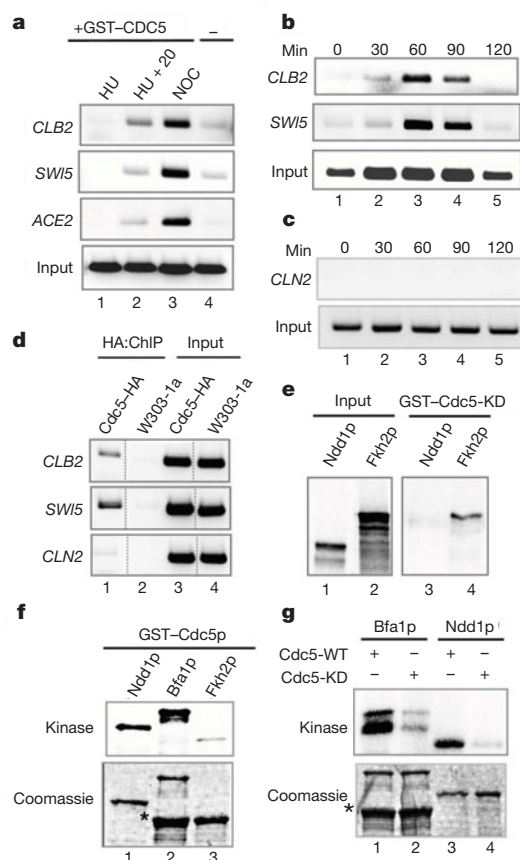


Figure 1 | Cdc5p is recruited to *CLB2* gene cluster promoters and phosphorylates Ndd1p. **a**, ChIP analysis with an anti-GST antibody and extracts isolated from mid-exponential-phase W303-1a cells (lane 4) or from MGY96 cells (lanes 1–3) treated with hydroxyurea (HU), released from hydroxyurea for 20 min (HU + 20), or treated with nocodazole (Noc). Input represents precipitated DNA amplified with *CLB2*-specific primers. **b**, **c**, ChIP analysis with an anti-GST antibody and extracts of MGY96 cells treated with α factor and released for the indicated times. Input is shown for the *CLB2* (**b**) and *CLN2* (**c**) promoters. **d**, ChIP analysis with anti-HA antibody and extracts from nocodazole-treated cells from either RB15 (expressing HA-tagged Cdc5p) or W303-1a strains. **e**, GST pull-down analysis of GST–Cdc5-KD and full-length Ndd1p and Fkh2p translated *in vitro*. A 25% input is shown. **f**, **g**, Protein kinase assays with wild-type (WT) GST–Cdc5p and full-length GST–Ndd1p, full-length MBP–Bfa1p or GST–Fkh2(458–862) as substrates (**f**) or WT or catalytically inactive (KD) GST–Cdc5p and full-length GST–Ndd1p or MBP–Bfa1p as substrates (**g**). The asterisk on the Coomassie gels represents a degradation product.

¹Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK. ²Institute for Cell and Molecular Biosciences, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. ³Division of Yeast Genetics, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. [†]Present address: EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, Final W5 Norte, Brasília, DF CEP 70770-900, Brazil.

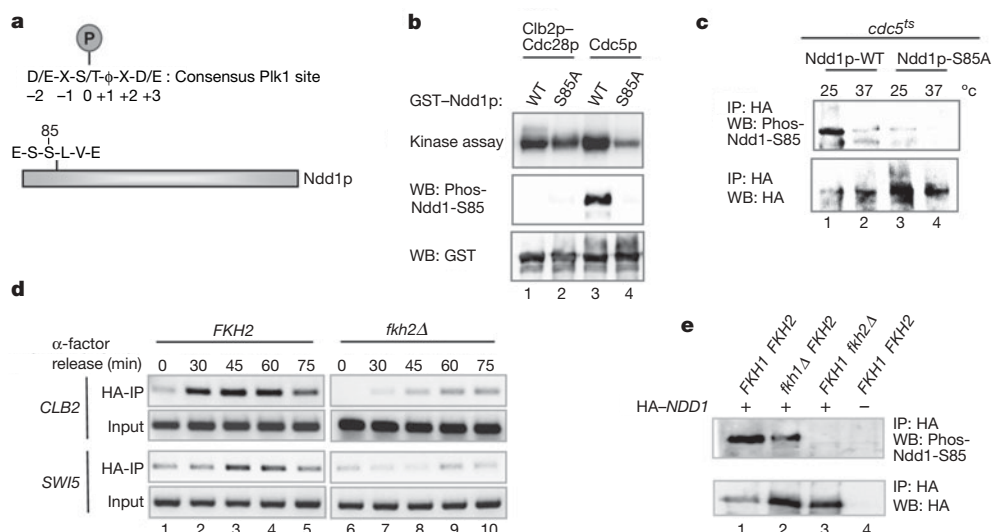


Figure 2 | Ndd1p is phosphorylated at Ser 85 by Cdc5p. **a**, The Cdc5p phosphorylation site, Ser 85, in Ndd1p and the consensus Plk site¹⁸. **b**, *In vitro* phosphorylation of wild-type (WT) and S85A mutant GST–Ndd1p by Clb2p–Cdc28p or Cdc5p was detected by either [³²P]phosphate incorporation (top) or western blot analysis (WB) with an anti-phospho-Ser 85 antibody (middle). Input protein was detected with an anti-GST antibody (bottom). **c**, *In vivo* phosphorylation of Ser 85 in Ndd1p. HA-tagged wild-type Ndd1p (WT) or Ndd1p(S85A) proteins were expressed in *cdc5^{ts}* cells (KKY021) grown at the indicated temperatures. IP, immunoprecipitation. **d**, ChIP analysis with anti-HA antibody and extracts

isolated from either RB15 (*FKH2*) or RB16 (*fkh2Δ*) strains expressing HA-tagged Cdc5p. Cells were taken at the indicated times after release from α -factor block. **e**, Fkh2p is required for phosphorylation of Ser 85 in Ndd1p. Wild-type HA-tagged Ndd1p was expressed in W303-1a (*FKH1 FKH2*), *fkh1Δ* or *fkh2Δ* cells (lanes 1–3). Lane 4 represents W303-1a (*FKH1 FKH2*) cells containing no HA-Ndd1p. In **c** and **e**, Ndd1p proteins were immunoprecipitated with anti-HA antibodies. Phosphorylation was detected by western blot analysis with an anti-phospho-Ser 85 antibody (top) and total precipitated protein was detected with an anti-HA antibody (bottom).

Cdc5p is difficult to detect because of its cyclical expression¹⁷. Although little Cdc5p could be detected in cells arrested in S phase by treatment with hydroxyurea, on release from the hydroxyurea block, recruitment of Cdc5p to the promoters of all of the *CLB2* cluster genes tested was observed, which was further enhanced when cells were arrested in M phase with nocodazole (Fig. 1a). Furthermore, after release from α -factor arrest in G1 phase, recruitment of Cdc5p was induced in a cell-cycle-dependent manner at both the *CLB2* and *SWI5* promoters, at a time consistent with the onset of transcription of these genes (Fig. 1b). However, Cdc5p recruitment to the promoter of a G1-phase-expressed cyclin-encoding gene, *CLN2* (ref. 13), could not be observed (Fig. 1c). Glutathione S-transferase (GST)-conjugated Cdc5p was expressed at times when promoter recruitment was not observed, demonstrating that recruitment is a temporal event that is not dependent on the levels of Cdc5p (Supplementary Fig. S1). ChIP analysis of a control strain expressing GST alone showed little enrichment of the *CLB2* promoter (Supplementary Fig. S2). Recruitment of Cdc5p expressed at endogenous levels could also be specifically detected at *CLB2* cluster promoters (Fig. 1d). Thus, despite the catalogued localization of Cdc5p to specific subcellular structures such as the spindle pole bodies² during mitosis, a portion of the cellular pool is associated with *CLB2* cluster promoters.

Likely candidates for recruiting Cdc5p to *CLB2* cluster promoters are components of the Mcm1p–Fkh2p–Ndd1p transcription factor complex. Indeed, Fkh2p, but not Ndd1p, formed a complex with Cdc5p *in vitro* (Fig. 1e). Similarly, GST–Cdc5p could co-precipitate Fkh2p when co-expressed *in vivo* (data not shown). The observed recruitment of Cdc5p to *CLB2* cluster promoters indicated that Cdc5p might directly phosphorylate one or more components of the Mcm1p–Fkh2p–Ndd1p complex. In contrast with Fkh2p, which was only poorly phosphorylated by Cdc5p, Ndd1p was phosphorylated as efficiently as the known Cdc5p substrate Bfa1p (Fig. 1f, and Supplementary Fig. S3)^{4,6}. A catalytically inactive version, Cdc5-KD, had a much lower ability to phosphorylate Ndd1p (Fig. 1g).

Ndd1p contains a perfect match at Ser 85 to the consensus phospho-acceptor motif for the mammalian polo kinase, Plk1 (ref. 18) (Fig. 2a).

We therefore replaced Ser 85 of Ndd1p with an alanine residue (S85A) and analysed the phosphorylation of the mutant protein by Cdc5p. *In vitro* kinase assays revealed a decrease in total phosphorylation of Ndd1p(S85A) by Cdc5p (Fig. 2b); moreover, western blotting with

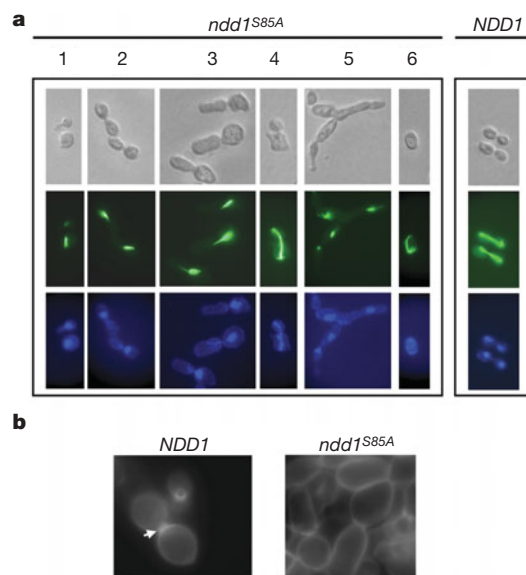


Figure 3 | Ser 85 of Ndd1p is required for several cell-cycle-dependent processes. Selection with 5-fluoro-orotic acid was used to replace wild-type Ndd1p in AP179 with *CEN*-based plasmids expressing the indicated wild-type and mutant Ndd1p proteins. **a**, Cellular morphologies were analysed by differential interference contrast microscopy (top). Mitotic spindles (middle) and nuclei (bottom) were analysed by fluorescence microscopy (spindles were revealed by fluorescein isothiocyanate, and nuclei by 4,6-diamidino-2-phenylindole). The numbers 1–6 refer to different cell morphologies (see Supplementary Information for details). **b**, Septa were revealed by fluorescence microscopy after treatment with calcofluor. The white arrow indicates normal septa.

an anti-phospho-Ser85 antibody showed that phosphorylation at Ser85 was abolished in Ndd1p(S85A) (Fig. 2b). In contrast, the S85A mutation had little effect on Ndd1p phosphorylation by Clb2p–Cdc28p complexes and phosphorylation at Ser85 was not observed (Fig. 2b).

To probe the potential role of Cdc5p in the phosphorylation of Ndd1p *in vivo*, we analysed the phosphorylation of wild-type and mutant (S85A) versions of Ndd1p in yeast strains containing a temperature-sensitive allele of *CDC5* (*cdc5^{ts}*). Significantly, phosphorylation of Ser85 could be detected in the wild-type, but not the mutant, Ndd1p at the permissive temperature (Fig. 2c). Furthermore, phosphorylation of Ser85 in wild-type Ndd1p was attenuated on inactivation of Cdc5p at 37 °C (Fig. 2c).

Clb2p–Cdc28p-mediated phosphorylation of Thr319 in Ndd1p is important in Ndd1p-mediated activation of the *CLB2* cluster¹⁶. However, phosphorylation of Thr319 is not required for Cdc5p-dependent phosphorylation of Ndd1p (Supplementary Fig. S4).

Fkh2p is important for the recruitment of Ndd1p to promoters of genes in the *CLB2* cluster⁹. As Cdc5p is also recruited to these promoters, and binds to Fkh2p *in vitro*, we tested whether Fkh2p is required for Cdc5p recruitment *in vivo*. Haemagglutinin (HA)-epitope-tagged Cdc5p was expressed in *FKH2* (wild-type control) and *fkh2Δ* strains, and promoter recruitment was monitored by ChIP

analysis. Cell-cycle-dependent temporal recruitment of Cdc5p to *CLB2* cluster promoters was observed in wild-type cells but was largely lost in *fkh2Δ* cells (Fig. 2d, and Supplementary Fig. S5). Using the same approach, we also determined whether Fkh2p was required for Ndd1p phosphorylation *in vivo*. Ndd1p was phosphorylated at Ser85 in wild-type and *fkh1Δ* cells, whereas phosphorylation of Ser85 was severely diminished in *fkh2Δ* cells (Fig. 2e). These results therefore suggest that Fkh2p has a function in nucleating the formation of a Fkh2p–Ndd1p–Cdc5p complex on *CLB2* cluster promoters, and the subsequent Cdc5p-dependent phosphorylation of Ndd1p at Ser85.

To understand the functional consequences of Cdc5p-mediated phosphorylation of Ndd1p at Ser85, we analysed the cellular phenotypes associated with mutations of this site. In contrast with wild-type control cells (more than 90% normal morphology), more than 90% of the *ndd1Δ* cells expressing Ndd1p(S85A) had aberrant morphologies, such as elongated daughter cells, hyphal-like growth and cells with problems in separating (Fig. 3a, and Supplementary Fig. S6). Moreover, these cells exhibited multiple different severe defects in nuclear morphology and mitotic spindle structure (Fig. 3a; see Supplementary Information for details). In addition, cytokinesis was defective, as demonstrated by the lack of formation of normal septa (Fig. 3b). Many of the defects observed in Ndd1p(S85A)-expressing cells resemble those observed in *cdc5* mutant cells, including

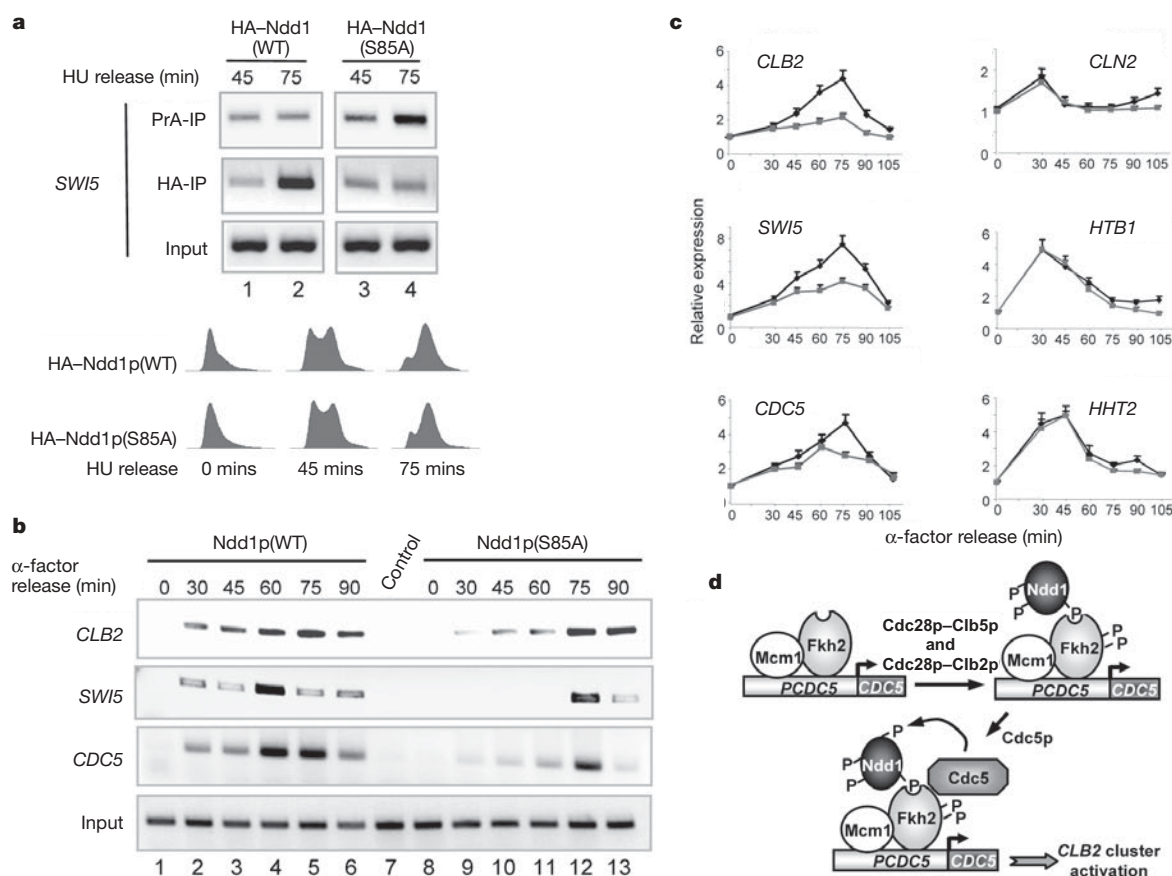


Figure 4 | Ser 85 of Ndd1p is required for normal promoter recruitment and *CLB2* gene cluster expression. **a**, ChIP analysis on the *SWI5* promoter was performed with either protein A (to bind TAP-tagged Ndd1p) or anti-HA antibody (to bind HA-tagged Ndd1p) and extracts isolated from S288C cells expressing TAP-tagged wild-type Ndd1p from the normal chromosomal locus and one of the indicated plasmid-borne HA-epitope-tagged Ndd1p proteins. Cells were synchronized with hydroxyurea (HU) and released for the indicated times. DNA content analysis of propidium iodide-stained cells expressing the indicated wild-type (WT) and mutant (S85A) Ndd1p proteins is shown. **b**, ChIP analysis on the *CLB2*, *SWI5* and *CDC5* promoters was performed with anti-HA antibody and extracts isolated from α -factor-

synchronized W303-1a cells expressing the indicated HA-epitope-tagged Ndd1p proteins. Times after release from α -factor arrest are indicated. Input is from the *CLB2* promoter. Control represents extracts from cells that do not express HA-Ndd1p. **c**, Real-time RT-PCR analysis of the indicated genes in the cells expressing only wild-type (WT; black diamonds) or mutant (S85A; grey squares) Ndd1p proteins (see Supplementary Figs. S9 and S10). Data are presented relative to the expression of each gene at the zero time point (taken as 1.0) and are shown as means $\pm$ s.d. **d**, Model of the effect of cell-cycle-dependent kinases on the regulation of the Mcm1p-Fkh2p-Ndd1p complex and the regulation of the *CLB2* cluster genes such as *CDC5*.

defects in septum formation and cytokinesis¹⁹. This indicates that at least some of the *cdc5*-dependent defects might be primarily due to deficiencies in Ndd1p-mediated gene regulation rather than problems in other functions of Cdc5p.

Next we examined whether phosphorylation of Ser 85 is involved in the recruitment of Ndd1p to promoters in the *CLB2* cluster. HA-epitope-tagged versions of Ndd1p were expressed from plasmids in a strain expressing Tandem Affinity Purification (TAP)-tagged wild-type Ndd1p from the normal chromosomal locus, to allow the cell cycle to proceed normally. Under these conditions, plasmid-expressed HA-tagged wild-type Ndd1p is more efficiently recruited to the *SWI5* promoter than HA-tagged Ndd1p(S85A) in the presence of TAP-tagged wild-type Ndd1p (Fig. 4a). In contrast, TAP-tagged wild-type Ndd1p is preferentially recruited only in the presence of plasmid-expressed HA-tagged Ndd1p(S85A) (Fig. 4a). These data indicated that wild-type Ndd1p might be more efficiently recruited to *CLB2* gene cluster promoters than Ndd1p(S85A). Indeed, consistent with this hypothesis was our observation of delayed recruitment of mutant Ndd1p(S85A) to several *CLB2* gene cluster promoters in synchronized cells co-expressing wild-type Ndd1p (Fig. 4b, and Supplementary Fig. S7). Moreover, fusion proteins containing Ndd1p and the DNA-binding domain of Gal4p (Gal4p-DBD) demonstrated equal recruitment of wild-type and mutant Ndd1p to a Gal4p-driven promoter but decreased recruitment of Ndd1p mutated at Ser85 to the *CLB2* promoter (Supplementary Fig. S8b). Taken together, these data therefore suggest that Cdc5p-mediated phosphorylation of Ndd1p at Ser85 is important in controlling the temporal recruitment of Ndd1p at promoters of *CLB2* cluster genes.

As a result of the severe defects observed in strains expressing only Ndd1p(S85A) (see Fig. 3), we were unable to analyse cell-cycle-dependent expression of the *CLB2* gene cluster in these cells. Therefore, to examine the potential role of Ser85 on *CLB2* cluster gene expression, we developed a system in which we could rapidly switch between endogenous and plasmid-borne expression of Ndd1p (Supplementary Fig. S9). The expression of *CLB2* cluster genes was then compared in cells expressing either wild-type Ndd1p or Ndd1p(S85A) after release from an α -factor block (Supplementary Fig. S10). Significantly, periodic gene expression that occurs early in the cell cycle in G1 phase (*CLN2*) and S phase (*HHT2* and *HTB1*) was similar in cells expressing either wild-type Ndd1p or Ndd1p(S85A) (Fig. 4c). In contrast, cells expressing Ndd1p(S85A) had decreased expression of the *CLB2* cluster genes *CLB2*, *SWI5* and *CDC5* (Fig. 4c). To investigate further the role of Cdc5p in controlling late cell-cycle-dependent gene expression, *cdc5^{ts}* cells were arrested in G1 phase, then released at the permissive (25 °C) or non-permissive (37 °C) temperature. In concordance with the effects observed in Fig. 4c, cell-cycle-dependent induction of *CLB2* was found to be severely attenuated on inactivation of Cdc5p (Supplementary Fig. S11). An additional role for phosphorylation of Ser85 in regulating the intrinsic transactivation function of Ndd1p was revealed with a *GAL1_{pr}-LacZ* reporter gene assay because Gal4p-DBD-Ndd1p(S85A) had a decreased ability to activate transcription (Supplementary Fig. S8a).

Taken together, our data show that Cdc5p-mediated phosphorylation of Ndd1p is important in controlling *CLB2* cluster gene expression and normal cell cycle progression. Because *CDC5* itself is part of the *CLB2* gene cluster, the phenotypic defects associated with mutation of Ser85 of Ndd1p are likely to be due in part to downstream defects caused by lower Cdc5p levels. Moreover, our data extend a previous model in which Clb2p-mediated phosphorylation of Ndd1p helps to establish a positive feedback loop that in turn leads to further increases in *CLB2* transcription and entry into M phase^{3,15,16}. Here Cdc5p is identified as another cell cycle regulator that can generate a positive feedback loop, because the *CDC5* gene is also part of the *CLB2* gene cluster, which is a target of the Mcm1p-Fkh2p-Ndd1p complex^{10,12,20–22}. Hence, Cdc5p and Clb2p-Cdc28p together con-

verge on Ndd1p to activate one of the key switches that controls cell cycle progression (Fig. 4d). This is reminiscent of the convergence of Cdc5p and Clb2p-Cdc28p on the regulation of Swe1p kinase and the subsequent promotion of mitotic entry²³. Thus, the combinatorial action of these kinases seems to be a common mechanism to control late cell cycle events precisely.

Forkhead transcription factors are also important for late cell-cycle-dependent gene expression in the evolutionarily divergent yeast *Schizosaccharomyces pombe*^{24–27} and in mammalian cells^{28,29}. Similarly, some of the previously described functions of polo kinases are also conserved between yeast and higher eukaryotes², and genetic studies link the Cdc5p homologue Plp1p to the regulation of M/G1-phase-specific transcription in *S. pombe*³⁰. Given these parallels, it is quite possible that a function for polo kinases in regulating cell-cycle-dependent gene expression in conjunction with forkhead transcription factor complexes will eventually be uncovered in mammalian systems.

METHODS

More detailed methods are supplied in Supplementary Information.

Plasmid constructs and yeast strains. Details of plasmid construction and yeast strains used in this study can be found in Supplementary Information.

RNA analysis and chromatin immunoprecipitations. RNA levels were analysed by northern blotting¹¹ or by real-time polymerase chain reaction with reverse transcription (RT-PCR; for further details see Supplementary Information). ChIP assays were performed as described previously¹⁴.

Protein production, pulldown assays and western blotting. GST pulldown assays with proteins translated *in vitro* were performed as described previously¹⁴. Fusion proteins with maltose-binding protein (MBP) were prepared by maltose-affinity chromatography (New England Biolabs). Western blots were performed with antibodies recognizing epitope tags or with an antibody raised against a phosphorylated peptide (NSSSNES[P]LVENS; Eurogentec).

Protein kinase production and kinase assays. Protein kinase assays using Flag-tagged Clb2p-kinase complexes purified from yeast cells, bacterially expressed and purified Clb2p and Cdc28p¹⁵ or GST-Cdc5p-WT and GST-Cdc5p-KD⁶ with bacterially expressed and purified substrates were performed as described previously¹⁵.

Received 13 July 2006; accepted 10 October 2006.

1. Barr, F. A., Silljé, H. H. W. & Nigg, E. A. Polo-like kinases and the orchestration of cell division. *Nature Rev. Mol. Cell Biol.* **5**, 429–441 (2004).
2. Lee, K. S., Park, J. E., Asano, S. & Park, C. J. Yeast polo-like kinases: functionally conserved multitask mitotic regulators. *Oncogene* **24**, 217–229 (2005).
3. Wittenberg, C. & Reed, S. I. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* **24**, 2746–2755 (2005).
4. Alexandru, G., Uhlmann, F., Mechtler, K., Poupard, M. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459–472 (2001).
5. Hu, F. *et al.* Regulation of the Bub2/Bfa1 GAP complex by Cdc5 and cell cycle checkpoints. *Cell* **107**, 655–665 (2001).
6. Geymonat, M., Spanos, A., Walker, P. A., Johnston, L. H. & Sedgwick, S. G. *In vitro* regulation of budding yeast Bfa1/Bub2 GAP activity by Cdc5. *J. Biol. Chem.* **278**, 14591–14594 (2003).
7. Sakchaisri, K. *et al.* Coupling morphogenesis to mitotic entry. *Proc. Natl Acad. Sci. USA* **101**, 4124–4129 (2004).
8. Fletcher, B. Transcriptional regulatory networks and the yeast cell cycle. *Curr. Opin. Cell Biol.* **14**, 676–683 (2002).
9. Koranda, M., Schleiffer, A., Endler, L. & Ammerer, G. Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters. *Nature* **406**, 94–98 (2000).
10. Kumar, R. *et al.* Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. *Curr. Biol.* **10**, 896–906 (2000).
11. Pic, A. *et al.* The forkhead protein Fkh2 is a component of the yeast cell cycle transcription factor SFF. *EMBO J.* **19**, 3750–3761 (2000).
12. Zhu, G. *et al.* Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* **406**, 90–94 (2000).
13. Breeden, L. L. Cyclin transcription: Timing is everything. *Curr. Biol.* **10**, R586–R588 (2000).
14. Pic-Taylor, A., Darieva, Z., Morgan, B. A. & Sharrocks, A. D. Regulation of cell cycle-specific gene expression through cyclin-dependent kinase-mediated phosphorylation of the forkhead transcription factor Fkh2p. *Mol. Cell Biol.* **24**, 10036–10046 (2004).

15. Darieva, Z. *et al.* Cell cycle regulated transcription through the FHA domain of Fkh2p and the coactivator Ndd1p. *Curr. Biol.* **13**, 1740–1745 (2003).
 16. Reynolds, D. *et al.* Recruitment of Thr 319-phosphorylated Ndd1p to the FHA domain of Fkh2p requires Clb kinase activity: a mechanism for CLB cluster gene activation. *Genes Dev.* **17**, 1789–1802 (2003).
 17. Cheng, L., Hunke, L. & Hardy, C. F. Cell cycle regulation of the *Saccharomyces cerevisiae* polo-like kinase *cdc5p*. *Mol. Cell. Biol.* **18**, 7360–7370 (1998).
 18. Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E. & Nishida, E. Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *J. Biol. Chem.* **278**, 25277–25280 (2003).
 19. Song, S. & Lee, K. S. A novel function of *Saccharomyces cerevisiae* CDC5 in cytokinesis. *J. Cell Biol.* **152**, 451–469 (2001).
 20. Althoefer, H., Schleiffer, A., Wassmann, K., Nordheim, A. & Ammerer, G. Mcm1 is required to coordinate G2-specific transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 5917–5928 (1995).
 21. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
 22. Simon, I. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708 (2001).
 23. Asano, S. *et al.* Concerted mechanism of Swe1/Wee1 regulation by multiple kinases in budding yeast. *EMBO J.* **24**, 2194–2204 (2005).
 24. Zilahi, E., Salimova, E., Simanis, V. & Sipiczki, M. The *S. pombe sep1* gene encodes a nuclear protein that is required for periodic expression of the *cdc15* gene. *FEBS Lett.* **481**, 105–108 (2000).
 25. Buck, V. *et al.* Fkh2p and Sep1p regulate mitotic gene transcription in fission yeast. *J. Cell Sci.* **117**, 5623–5632 (2004).
 26. Bulmer, R. *et al.* The forkhead transcription factor Fkh2 regulates the cell division cycle of *Schizosaccharomyces pombe*. *Eukaryot. Cell* **3**, 944–954 (2004).
 27. Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nature Genet.* **36**, 809–817 (2004).
 28. Alvarez, B., Martinez, A. C., Burgering, B. M. & Carrera, A. C. Forkhead transcription factors contribute to execution of the mitotic programme in mammals. *Nature* **413**, 744–747 (2001).
 29. Laoukili, J. *et al.* FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nature Cell Biol.* **7**, 126–136 (2005).
 30. Anderson, M. *et al.* Plo1⁺ regulates gene transcription at the M–G₁ interval during the fission yeast mitotic cell cycle. *EMBO J.* **21**, 5745–5755 (2002).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank A. Clancy for technical assistance; I. Hagan, P. March, M. Jackson and G. Pereira for advice; A. Whitmarsh, S.-H. Yang and members of our laboratories for comments on the manuscript and helpful discussions; and M. Walberg, U. Surana, G. Pereira and D. Lydall for reagents. This work was supported by Cancer Research UK, the BBSRC, the MRC and the Wellcome Trust.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.D.S. (a.d.sharrocks@man.ac.uk) or B.A.M. (b.a.morgan@ncl.ac.uk).

In vivo enhancer analysis of human conserved non-coding sequences

Len A. Pennacchio^{1,2}, Nadav Ahituv², Alan M. Moses², Shyam Prabhakar², Marcelo A. Nobrega^{2,†}, Malak Shoukry², Simon Minovitsky², Inna Dubchak^{1,2}, Amy Holt², Keith D. Lewis², Ingrid Plajzer-Frick², Jennifer Akiyama², Sarah De Val⁴, Veena Afzal², Brian L. Black⁴, Olivier Couronne^{1,2}, Michael B. Eisen^{2,3}, Axel Visel² & Edward M. Rubin^{1,2}

Identifying the sequences that direct the spatial and temporal expression of genes and defining their function *in vivo* remains a significant challenge in the annotation of vertebrate genomes. One major obstacle is the lack of experimentally validated training sets. In this study, we made use of extreme evolutionary sequence conservation as a filter to identify putative gene regulatory elements, and characterized the *in vivo* enhancer activity of a large group of non-coding elements in the human genome that are conserved in human–pufferfish, *Takifugu (Fugu) rubripes*, or ultra-conserved¹ in human–mouse–rat. We tested 167 of these extremely conserved sequences in a transgenic mouse enhancer assay. Here we report that 45% of these sequences functioned reproducibly as tissue-specific enhancers of gene expression at embryonic day 11.5. While directing expression in a broad range of anatomical structures in the embryo, the majority of the 75 enhancers directed expression to various regions of the developing nervous system. We identified sequence signatures enriched in a subset of these elements that targeted forebrain expression, and used these features to rank all 3,100 non-coding elements in the human genome that are conserved between human and *Fugu*. The testing of the top predictions in transgenic mice resulted in a threefold enrichment for sequences with forebrain enhancer activity. These data dramatically expand the catalogue of human gene enhancers that have been characterized *in vivo*, and illustrate the utility of such training sets for a variety of biological applications, including decoding the regulatory vocabulary of the human genome.

Significant progress has been made in the identification of core promoter elements based on their defined position immediately upstream of each gene and their nearly universal activation by RNA polymerase II^{2,3}. However, the identification of distant acting gene regulatory sequences that direct precise spatial and temporal patterns of expression has been limited, despite their established roles in development⁴, phenotypic diversity⁵ and human disease^{6–8}. Comparative genomic-based approaches have proved to be useful in identifying gene regulatory sequences, primarily on a gene-by-gene basis. These studies involved sequence comparisons of human (or other vertebrate) genomic intervals to orthologous regions from organisms separated by varying evolutionary distances, ranging from primates to fish^{9–12}. From this work it has been implied that ancient conservation (such as between human and fish) as well as ‘ultra’-conservation among mammals (sequences at least 200 base pairs (bp) in length that are 100% identical among human/mouse/rat)¹ may be useful indicators of sequences with an increased likelihood of demonstrating gene regulatory activity. These gene-centric investigations,

however, have identified only a relatively small number of distant-acting enhancer sequences.

As one of the goals of this work was to assess the validity of a genome-based approach, rather than a gene-centric one, we chose non-coding target sequences based on one of two ‘extreme’ comparative genomic criteria: ancient conservation between human and *Fugu* (separated by ~450 million years of evolution) or ultra-conservation among human/mouse/rat¹. In total, 167 human DNA fragments were assessed for spatial enhancer activity in a well-established transgenic mouse enhancer assay that links the human conserved fragment to a minimal mouse heat shock promoter fused to a *lacZ* reporter gene^{10,13–16}. We chose to determine tissue-specific reporter gene expression at embryonic day 11.5 (e11.5), as this developmental stage allows for whole-mount staining and whole-embryo visualization. Moreover, at this time-point many of the major tissues and organs have been specified. We also expected this stage to be particularly informative because ‘extreme’ conserved non-coding elements tend to be enriched and clustered near genes expressed during embryonic development^{1,12,17,18}.

Overall, we found that 29% (24/83) of human–*Fugu* elements alone and 61% (33/54) of human–*Fugu* elements that are also ultraconserved were positive enhancers in this *in vivo* assay

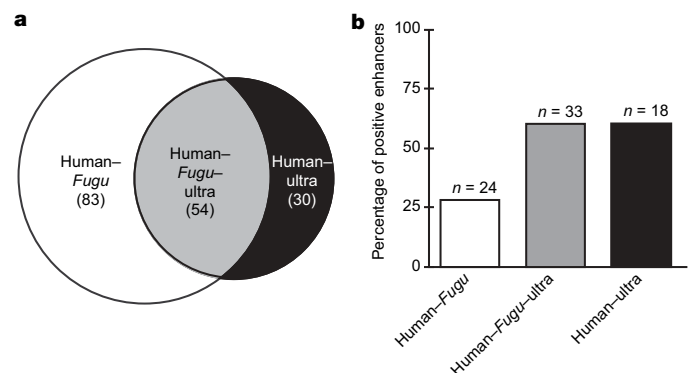


Figure 1 | A summary of all sequences tested for enhancer activity in transgenic mice. a, A breakdown of the assayed non-coding sequences by human–*Fugu* conservation and/or human–rodent ultraconservation: Human–*Fugu* only, human–*Fugu* and human–rodent, or human–rodent only. **b**, The total percentage of positive human enhancers broken down by the same parameters as described in **a**. The total number of elements tested is indicated within **a**, while the number of positives is found above the bars of the graph in **b**.

¹US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. ²Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ³Molecular and Cellular Biology Department, University of California-Berkeley, California 954720, USA. ⁴Cardiovascular Research Institute, University of California, San Francisco, California 94143-2240, USA. [†]Present address: Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.

(Fig. 1; Supplementary Table 1; the entire data set including the sequence coordinates, conservation, and whole-mount embryo digital imagery can be accessed and queried at the VISTA Enhancer Browser, <http://enhancer.lbl.gov>). As an example of these data, we present 23 elements meeting our selection criteria that were located in a gene-poor 2.5 Mb stretch bracketing *SALL1*, a gene encoding a transcription factor expressed in early development and mutated in Townes-Brooks syndrome¹⁹ (Fig. 2). Seven of the elements flanking *SALL1* directed tissue-specific reporter gene expression in the transgenic *in vivo* assay, recapitulating aspects of *SALL1*'s endogenous expression characteristics at e11.5²⁰ and further supporting the postulated modular nature of distant acting gene enhancers^{21,22}. In addition, we tested 30 ultraconserved non-coding sequences that lacked identifiable conservation with *Fugu* of which 18 (60%) functioned as enhancers, similar to the success rate observed for ultraconserved elements that also have *Fugu* conservation (Fig. 1). Whereas the average size of the human fragments tested was 1,270 bp, the positive enhancers overlapped longer human–rodent conserved regions (average length 1,630 bp versus 966 bp; *t*-test *P*-value=0.0087; see Supplementary Methods) and were more conserved among mammals (human–rodent conservation score, *t*-test *P*-value=0.0004; see Supplementary Methods) relative to negatives in the assay.

These experimental results reveal the high propensity of extremely conserved human non-coding sequences to behave as transcriptional enhancers *in vivo*, and support both ancient human–fish conservation and human–rodent ultraconservation as highly effective filters to identify such functional elements. The large percentage of elements positive for enhancer activity is particularly surprising, considering the single time-point of investigation and the likely possibility that a fraction of the negatives may be enhancers active either earlier or later in development. An important question arising from the significant fraction of ultra and *Fugu* conserved elements functioning as enhancers is whether the tissue-specific enhancer activity that we assess completely explains why these sequences are so constrained. Overlaying our data set with results from a recent ChIP–Chip study²³ indicates that at least seven of the elements reported here (including four that are enhancers at e11.5) presumably function as gene silencers in embryonic stem cells. Such data imply that functions in addition to tissue-specific transcriptional activation are embedded in some fraction of extremely conserved

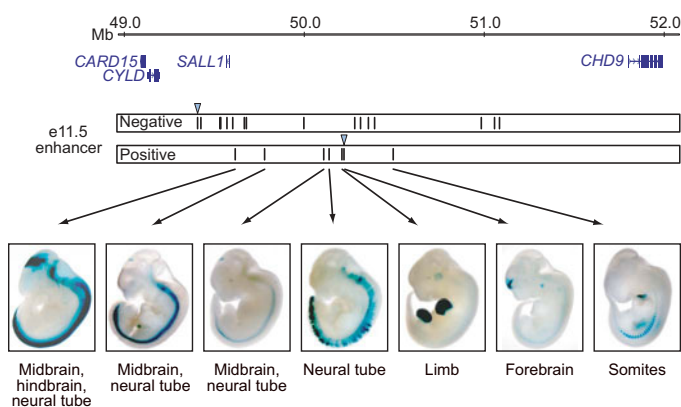


Figure 2 | A 3 Mb region of human chromosome 16 enriched for human–*Fugu* non-coding conservation flanking the *SALL1* gene. The coordinates and gene annotations located at the top of the diagram are based on the hg17 assembly at the UCSC Genome Browser (<http://genome.ucsc.edu>). The middle tracks depict human fragments that were tested in the transgenic mouse enhancer assay, and their classification as either 'negative' or 'positive' refers to their enhancer activity at e11.5. All human elements tested were conserved in the *Fugu* genome, and two of these elements were also defined as ultraconserved (denoted by arrowheads). The bottom panel indicates the positive enhancer activities captured through transgenic mouse testing of human–*Fugu* conserved non-coding fragments in this interval.

non-coding elements, thus potentially contributing to their extreme level of constraint. However, the high efficiency of enhancer identification through this approach nonetheless suggests that tissue-specific transcriptional enhancer activity may be one of the predominant functions of non-coding genomic regions under extreme constraint throughout vertebrate evolution.

We categorized all 75 identified enhancers by their general anatomical patterns of expression using an existing standardized nomenclature²⁴ (Fig. 3). All positive enhancer annotations are based on a minimum of three independent transgenic *F₀* embryos carrying the same construct and demonstrating the same expression pattern, though the majority (83%) had four or more supporting embryos. We observed reporter gene expression in a variety of anatomical regions, including embryonic structures that are subject to major morphogenetic and remodelling events at e11.5, such as the developing limb, the somites, the heart and the branchial arches (Fig. 3). Of the 16 distinct anatomical structures where expression was noted, it was most frequently observed in the central and peripheral nervous system, with the most prevalent patterns corresponding to forebrain, midbrain, neural tube, and hindbrain (Fig. 3). This bias may be partially explained by the intrinsic complexity of the genetic cascades underlying vertebrate nervous system development²⁵ as well as the high percentage of all genes that are expressed in the nervous system.

The majority of the enhancers (50 elements, 66%) directed reproducible expression only to a single anatomical structure at the resolution of whole-mounts. This is consistent with the notion that complex endogenous messenger RNA expression patterns commonly result from the combined effects of several independent *cis*-regulatory sequences. The remaining one-third (25/75) of the enhancers directed expression to two or more anatomical structures. We speculate that these enhancer elements may be composed of two or more adjacent functional modules that are too tightly linked to each other to be resolved by our comparative approach, or that several tissue-specific enhancer activities overlap within a single enhancer element that is used in more than one developmental process. Importantly, the enhancer data set reported here provides a sizeable sequence-based substrate to begin to dissect these possible regulatory mechanisms, as well as reagents for further in-depth biological investigation.

To explore if our *in vivo* enhancer data set could be used to identify sequence features associated with elements driving reporter gene expression in specific anatomical structures, we focused on the forebrain as a test case and selected as a training set four of the strongest

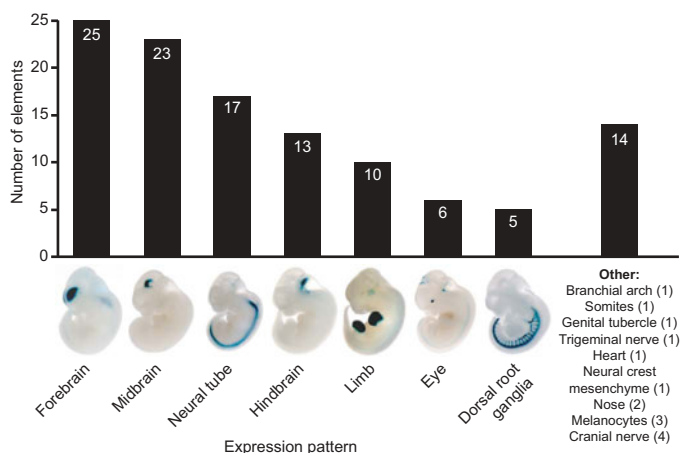


Figure 3 | Grouping of positive expression patterns captured in the transgenic mouse enhancer assay. The total number of elements displaying a given anatomical pattern is depicted by the height of the bars in the chart. A representative transgenic embryo is provided for each expression pattern. Elements with reproducible staining in more than one structure are included in each respective category.

enhancers identified early on in our survey. Using a motif finding strategy, we identified six motifs significantly over-represented in these enhancers (see Supplementary Methods). We then scored and ranked all ~3,100 human–*Fugu* conserved non-coding elements in the human genome (Supplementary Table 2) for the statistical over-representation of these putative forebrain motifs (Supplementary Methods). The 30 highest-ranking elements included the four known forebrain enhancers that constituted the training set as well as 26 additional elements (Supplementary Table 3). Of these 26 elements, 23 were successfully cloned and tested for *in vivo* enhancer activity in transgenic mice. We observed robust forebrain enhancer activity for 4 of the 23 elements (17%) tested in the transgenic assay system. By comparison, only 4 (5%) of the 77 otherwise uncharacterized human–*Fugu* conserved elements used to identify the training set were forebrain enhancers (see Supplementary Methods; Fig. 4). This preliminary result, although based on a small training set, indicates that a combined comparative and motif-based strategy provided a greater than threefold enrichment ($P = 0.08$, see Supplementary Methods) over comparative-only approaches for the identification of enhancers active in a particular tissue of interest. This initial computational investigation also highlights the need for larger characterized enhancer training sets, the annotation of tissue specificities at high spatial resolution and the development of improved computational methods, which will probably provide a substrate to conclusively establish the predictive power of such approaches.

This study provides quantitative support for previous anecdotal observations that ‘extreme’ evolutionary non-coding conservation is a powerful predictor of mammalian tissue-specific enhancers. Of note, there are at least an additional 5,500 human–fish conserved non-coding sequences in the human genome with similar levels of

constraint that are strong candidates for acting as gene enhancers¹¹. The efficiency of enhancer identification coupled with the relatively high throughput transgenic assay used here represents a feasible approach for the generation of a genome-wide experimentally validated enhancer data set. Such collections are expected to define functional candidate regions as medical sequencing efforts escalate, as well as provide a foundation for inferring the network of regulatory interactions among key developmental genes during vertebrate development, analogous to the well-developed efforts in non-vertebrate model systems^{21,22}. Regulatory insights derived from these analyses should also enable the creation of modules driving pre-determined expression patterns for various biological applications, as well as contribute to an understanding of the vocabulary and grammar of DNA sequences dictating gene expression.

METHODS

Identification of conserved elements and transgenic enhancer assay. Human–*Fugu* conserved non-coding elements with 70% identity, a score of match-mismatch ≥ 60 , and lacking evidence of encoding a protein or being transcribed in mRNA were derived from whole-genome alignments (see Supplementary Methods). The coordinates of ultraconserved elements were retrieved from ref. 1. Conserved elements were amplified from human genomic DNA by polymerase chain reaction (PCR), sequence-validated and transferred into an *Hsp68-LacZ* reporter vector. Generation of transgenic mice and embryo staining was done as previously described²⁶ in accordance with protocols approved by the Lawrence Berkeley National Laboratory. For each enhancer fragment, all transgenic embryos exhibiting LacZ-staining were scored and annotated independently by multiple curators.

Motif identification and prediction of forebrain enhancers. To find sequence motifs that were associated with forebrain expression, we used a discrete, enumerative motif-finding approach²⁷. We identified motifs enriched in the training set of forebrain enhancers relative to three sets of background sequences: (1) random sequences from chromosome 16 (ATTAA and GATTA, which we note are motifs present in previously characterized embryonic forebrain enhancers^{28,29}), (2) a chromosome 16 set of human–*Fugu* conserved elements (TTNNAAG, CANNNGC and TANNTGA), and (3) a chromosome 16 set of human–*Fugu* sequences that displayed enhancer activity (TTNNNTT) (see Supplementary Methods for details). We then combined information from all the motifs for the prediction of new forebrain enhancers in the genome by scoring each of 3,124 human/mouse/*Fugu* non-coding alignments¹⁸ for the number of conserved (found aligned in human/mouse/*Fugu*) matches to each of the 6 significant 5-mers (see Supplementary Methods for details of scoring procedure). The top 30 fragments are available in Supplementary Table 3.

Received 14 June; accepted 22 September 2006.

Published online 5 November 2006.

	Coordinates (genes)	Pattern	Positive
a Training set	Chr. 16: 77068109–77069445 (<i>WWOX</i> , intragenic)		7/10
	Chr. 16: 70812067–70813326 (<i>PMFBP1-ATBF1</i>)		6/7
	Chr. 16: 53208099–53209383 (<i>IRX3-IRX5</i>)		5/5
	Chr. 16: 50228682–50229540 (<i>SALL1-CHD9</i>)		4/6
b Predicted enhancers	Chr. 1: 87533642–87535103 (<i>LMO4-PKN2</i>)		9/12
	Chr. 3: 71373108–71375274 (<i>FOXP1</i> , intragenic)		8/9
	Chr. 3: 182256341–182258504 (<i>TIM14-SOX2</i>)		6/8
	Chr. X: 24675039–24677929 (<i>POLA</i> , intragenic)		5/6

Figure 4 | Application of a forebrain enhancer training set to identify forebrain-specific enhancer sequences elsewhere in the human genome. **a**, The four human–*Fugu* chromosome 16 forebrain enhancers used in the training set. **b**, The four positive forebrain enhancers from the 23 human–*Fugu* genome-wide elements predicted to direct forebrain expression based on the training set in **a**. The UCSC human genome coordinates of the tested fragments (May 2004) and the flanking gene(s) are provided, as well as three representative embryos. Numbers indicate the ratio of forebrain-positive to the total number of stained embryos.

- Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
- Roeder, R. G. & Rutter, W. J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234–237 (1969).
- Goldberg, M. L. *Sequence Analysis of Drosophila Histone Genes*. Ph.D. thesis, Stanford Univ. (1979).
- Stathopoulos, A. & Levine, M. Genomic regulatory networks and animal development. *Dev. Cell* **9**, 449–462 (2005).
- Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).
- Emison, E. S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857–863 (2005).
- Kleinjan, D. A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8–32 (2005).
- Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
- Boffelli, D. *et al.* Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391–1394 (2003).
- Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
- Prabhakar, S. *et al.* Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res.* **16** (7), 855–863 (2006).
- Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
- Kothary, R. *et al.* Inducible expression of an *hsp68-lacZ* hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).

14. Rojas, A. *et al.* Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development* **132**, 3405–3417 (2005).
15. Rossant, J., Zirngibl, R., Cado, D., Shago, M. & Giguere, V. Expression of a retinoic acid response element-hsplacZ transgene defines specific domains of transcriptional activity during mouse embryogenesis. *Genes Dev.* **5**, 1333–1344 (1991).
16. Yamagishi, H. *et al.* Tbx1 is regulated by tissue-specific forkhead proteins through a common Sonic hedgehog-responsive enhancer. *Genes Dev.* **17**, 269–281 (2003).
17. Boffelli, D., Nobrega, M. A. & Rubin, E. M. Comparative genomics at the vertebrate extremes. *Nature Rev. Genet.* **5**, 456–465 (2004).
18. Ahituv, N., Prabhakar, S., Poulin, F., Rubin, E. M. & Couronne, O. Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. *Hum. Mol. Genet.* **14**, 3057–3063 (2005).
19. Kohlhasse, J., Wischermann, A., Reichenbach, H., Froster, U. & Engel, W. Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nature Genet.* **18**, 81–83 (1998).
20. Buck, A., Kispert, A. & Kohlhasse, J. Embryonic expression of the murine homologue of SALL1, the gene mutated in Townes–Brocks syndrome. *Mech. Dev.* **104**, 143–146 (2001).
21. Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3**, e245 (2005).
22. Davidson, E. H. *Genomic Regulatory Systems: In Development and Evolution* (Academic, San Diego, 2001).
23. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
24. Bard, J. L. *et al.* An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mech. Dev.* **74**, 111–120 (1998).
25. Gray, P. A. *et al.* Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255–2257 (2004).
26. Poulin, F. *et al.* *In vivo* characterization of a vertebrate ultraconserved enhancer. *Genomics* **85**, 774–781 (2005).
27. van Helden, J., Andre, B. & Collado-Vides, J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* **281**, 827–842 (1998).
28. Kurokawa, D. *et al.* Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. *Development* **131**, 3319–3331 (2004).
29. Zhou, J., Zwicker, J., Szymanski, P., Levine, M. & Tjian, R. TAFII mutations disrupt Dorsal activation in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **95**, 13483–13488 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Research was conducted at the E. O. Lawrence Berkeley National Laboratory, under the Programs for Genomic Application, funded by the National Heart, Lung, and Blood Institute, USA as well as the National Human Genome Research Institute, USA, and performed under a Department of Energy Contract with the University of California.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.A.P. (LAPennacchio@lbl.gov).

Structure of the *E. coli* signal recognition particle bound to a translating ribosome

Christiane Schaffitzel¹, Miro Oswald¹, Imre Berger¹, Takashi Ishikawa¹, Jan Pieter Abrahams², Henk K. Koerten³, Roman I. Koning³ & Nenad Ban¹

The prokaryotic signal recognition particle (SRP) targets membrane proteins into the inner membrane^{1–4}. It binds translating ribosomes and screens the emerging nascent chain for a hydrophobic signal sequence, such as the transmembrane helix of inner membrane proteins. If such a sequence emerges, the SRP binds tightly, allowing the SRP receptor to lock on. This assembly delivers the ribosome–nascent chain complex to the protein translocation machinery in the membrane. Using cryo-electron microscopy and single-particle reconstruction, we obtained a 16 Å structure of the *Escherichia coli* SRP in complex with a translating *E. coli* ribosome containing a nascent chain with a transmembrane helix anchor. We also obtained structural information on the SRP bound to an empty *E. coli* ribosome. The latter might share characteristics with a scanning SRP complex, whereas the former represents the next step: the targeting complex ready for receptor binding. High-resolution structures of the bacterial ribosome and of the bacterial SRP components are available, and their fitting explains our electron microscopic density. The structures reveal the regions that are involved in complex formation, provide insight into the conformation of the SRP on the ribosome and indicate the conformational changes that accompany high-affinity SRP binding to ribosome nascent chain complexes upon recognition of the signal sequence.

The targeting of membrane proteins by SRPs is universal for all kingdoms of life^{1–4}, and eukaryotic SRPs (which contain a 300-nucleotide RNA and six proteins⁵) also target presecretory proteins to the endoplasmic reticulum. Prokaryotic SRPs are smaller; the *E. coli* SRP consists of a 4.5S RNA (Ffs) and a single 48-kDa protein (Ffh, fifty-four homologue). The Ffh protein has three domains, termed N, G and M. The N and G domains are structurally and functionally coupled¹. Ribosome binding is mediated through the N domain^{6,7}, which has a four-helix bundle. The G domain has a Ras-like GTPase fold with an additional insertion box motif that is unique to SRP GTPases. The methionine-rich M domain contains the hydrophobic signal sequence binding pocket and binds Ffs^{1,3,8}. Both Ffs and Ffh are conserved and are essential for viability, and hence they represent a minimal functional version of the complex. A cryo-electron microscopy (EM) structure of a eukaryotic SRP–ribosome complex is available⁹, but at present there are no structures of prokaryotic SRP–ribosome complexes.

Stable *E. coli* ribosomal nascent chain complexes (RNCs) were generated by *in vitro* translation¹⁰. The nascent chain contained the transmembrane helix (TMH) of *E. coli* FtsQ, which had been previously observed to cross-link to Ffh¹¹. After *in vitro* translation, the RNCs were purified by sucrose gradient centrifugation and affinity chromatography (Supplementary Fig. 1). The SRP bound strongly

even to non-translating ribosomes under physiological salt conditions. Only under high-salt conditions could a significantly higher affinity of SRPs to RNCs than to empty ribosomes be detected (Supplementary Fig. 1). Under the conditions chosen for the EM experiments, virtually all RNCs were in complex with SRPs, which is consistent with observed dissociation constants (~ 0.1 nM for the RNCs and 50–90 nM for 70S ribosomes)^{6,12,13}. The three-dimensional cryo-EM reconstruction of the RNC–SRP complex has a resolution of 16 Å (Fourier shell correlation (FSC) 0.5, see Supplementary Fig. 1; 11 Å according to the 3σ criterion), whereas that of the 70S–SRP complex is about 20 Å (see Supplementary Table 1).

In the RNC, all three active sites of the ribosome are occupied by transfer RNAs. However, the occupancy at the A-site is significantly lower than at the fully occupied P- and E-sites, so the nascent chain is connected mostly to P-site tRNA. The SRP is positioned at the polypeptide tunnel exit, as expected. The RNA part of the SRP can be recognized in the density by its helical appearance. The SRP adopts an elongated shape (170 Å × 75 Å) that projects away from the polypeptide tunnel exit (Fig. 1). The polypeptide tunnel exit is not covered by the SRP, and the nascent chain is still accessible to other factors. This is a marked difference from the eukaryotic SRP, which covers the polypeptide tunnel exit⁹. Another important difference in *E. coli* SRP is the absence of the Alu domain, which in eukaryotes stalls translation by interfering with elongation-factor binding⁹.

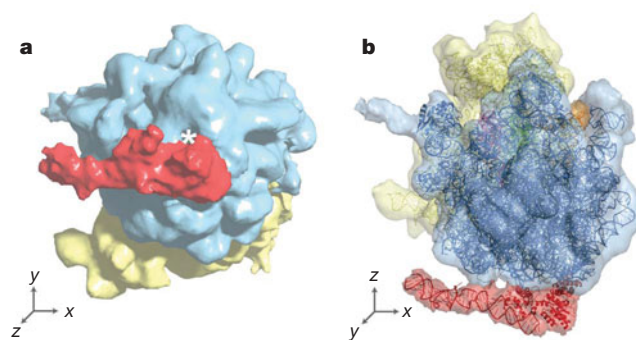


Figure 1 | Cryo-EM structure of the RNC–SRP complex. Ribosomal subunits are shown in yellow (30S) and blue (50S). The SRP is shown in red and is displayed at a lower contour level than the RNC complex. The RNC–SRP complex is shown with the view into the polypeptide exit tunnel (white star) (a) and from the back of the 50S subunit (b). A-, P- and E-site tRNAs are coloured magenta, green and orange, respectively. The crystal structures of 70S, tRNAs and a molecular model of SRP have been fitted into the density.

¹ETH Zurich, Institute for Molecular Biology and Biophysics, HPK Building, Schafmattstrasse 20, 8093 Zurich, Switzerland. ²Department of Biophysical Structural Chemistry, Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2300 RA Leiden, The Netherlands. ³Leiden University Medical Center, Molecular Cell Biology, Center for Electron Microscopy, PO Box 9600, 2300 RC Leiden, The Netherlands.

The coordinates of the *E. coli* 70S ribosome crystal structure¹⁴ were fitted into the density using SITUS¹⁵. The SRP was modelled as two rigid bodies from two sources: the NG domain of *Thermus aquaticus* in complex with the non-hydrolysable GTP analogue GMPPNP¹⁶; and the M domain in complex with RNA domain IV from *Sulfolobus solfataricus*¹⁷. Finally, the remainder of the 4.5S RNA was built as an extension of domain IV using a model from the SRP database¹⁸. The density revealed a bipartite structure for 4.5S RNA with two mostly straight helical regions of around 55 Å and 77 Å in length, and a pronounced kink of approximately 130° at the joint (Fig. 2). After manual readjusting, the SRP model was further improved using normal mode flexible fitting¹⁹ and energy minimization with CNS software²⁰.

Independent fitting of the NG and M domains in complex with the 4.5S RNA resulted in an SRP conformation that is different from the crystal structure of the archaeal SRP core¹⁷. Our data indicated a 180° rotation about a flexible hinge region at the carboxyl terminus of the G domain in addition to a 130° rotation along the axis of the linker helix between the NG and M domains of Ffh (Fig. 2c). These flexible regions stretch upon binding and position the M domain almost 20 Å away from the NG domain—resulting in a more elongated, less compact shape of bound as compared to unbound SRP. The crystal structure of the archaeal SRP core has been suggested to present a closed, compact conformation of the SRP, which could exist in the non-ribosome bound state and could rearrange and open up upon ribosome binding²¹.

We identified four distinct contact sites (c1–c4) between the SRP and the large ribosomal subunit (Fig. 3a), but in the absence of a signal peptide in the RNC, only the c1 contact remains (see Supplementary Fig. 3). In the first contact site (c1), the N domain of Ffh mainly contacts ribosomal protein L23, and to a far lesser degree L29 (Fig. 3a, b). This is in agreement with previous cross-linking experiments^{6,22}. C1 is conserved in the eukaryotic structure⁹, where flexible loops of the N domain contact L23 and L29, so that the

domain is positioned almost perpendicular to the surface of the ribosome (see Supplementary Fig. 4). In our structure, by contrast, the N domain contacts L23 with its helices and the G domain is in the vicinity of L29. Consequently, the relative positioning of the NG and M domains is different (see Supplementary Information).

The contact area between Ffh and L23/L29 is formed mainly by three helices (h1–h3) of the N domain. It features a patch of positively charged residues at the binding interface (see Supplementary Fig. 2) juxtaposed to the negatively charged residues on the surface of L23, which could explain the salt-sensitivity of the 70S–SRP complex (see Supplementary Fig. 1c). The G domain of Ffh is positioned above L29 and could also contribute to ribosome binding. In contrast to eukaryotic SRP⁷, we did not find crosslinks between the *E. coli* SRP and L29, perhaps because residues 17 and 25 used in the crosslinking studies are positioned at the tip of the N domain four-helix bundle and point away from L29 in our structure^{6,12}.

Strikingly, L23 is also the main contact point of two other factors that interact with nascent chains: the translocon and the ribosome-associated chaperone trigger factor (TF). At the last stage of co-translational protein targeting, the translocon, SecYEG, replaces the SRP and their binding is mutually exclusive. Consequently, we observe that the SRP and SecYEG¹⁰ use almost identical contact areas (c1, c2 and c4) for their interactions with the ribosome. On the other hand, biochemical evidence exists showing that the SRP and TF can bind simultaneously to the ribosome^{12,13} and compete for the nascent chain²³. In our structure, the contact between the SRP and L23, which is mediated by the N domain of the SRP, does not seem to overlap with the anchor point of TF, which interacts with a relatively small surface area of the ribosome using a flexible loop (see Supplementary Information)²⁴. However, an SRP bound to the translating ribosome as visualized in this study would sterically hinder access of TF to its ribosomal binding site. Therefore, an additional conformational change in SRP would be necessary to explain simultaneous binding of TF and SRP to the ribosome. It is possible that this is the reason for

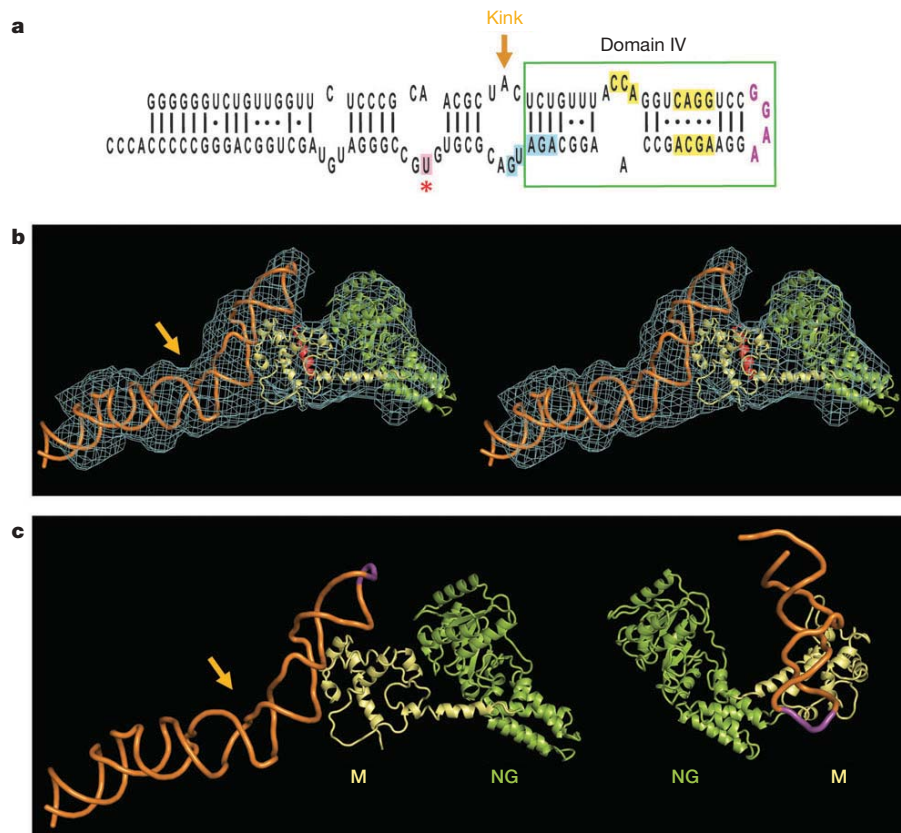


Figure 2 | Molecular model of the SRP bound to a translating ribosome. **a**, Secondary structure of SRP RNA with the M-domain binding site (yellow), ribosome contact region (blue), tetra-loop (magenta) and crosslink site U84 (red star). The conserved domain IV is marked with a green box. **b**, Stereo view of the molecular model of *E. coli* SRP. Density is shown in cyan, 4.5S RNA in orange, Ffh M domain in yellow, Ffh NG domain in green and the TMH in red. **c**, Comparison between the molecular model of *E. coli* SRP (left) and the crystal structure of SRP54 and SRP RNA domain IV¹⁷ (right). For orientation, the RNA tetra-loop is shown in magenta. The kink region is marked by an orange arrow.

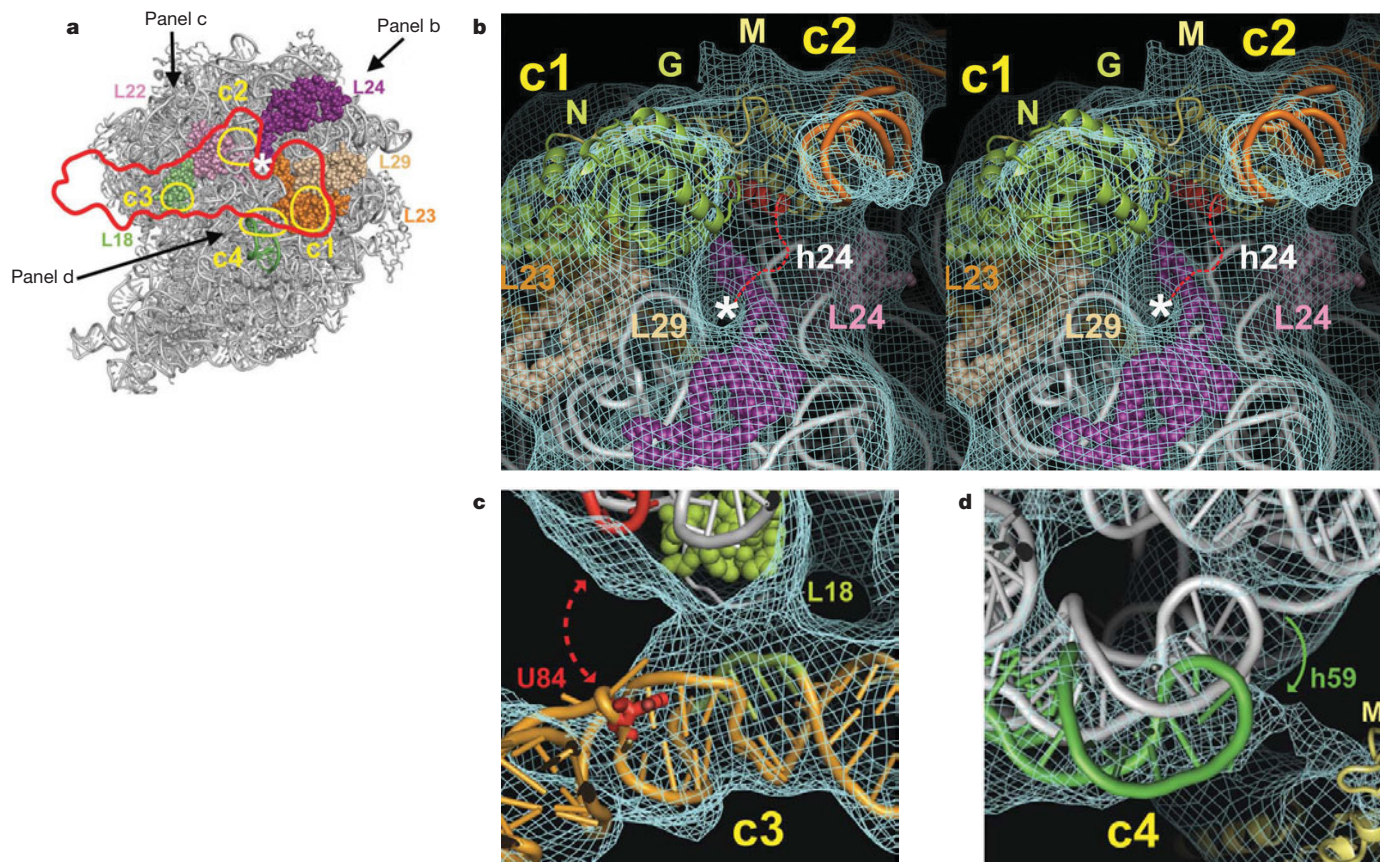


Figure 3 | Contact sites of SRP on 50S. **a**, Overview. SRP is outlined in red and the four contacts in yellow on the crystal structure of 70S. **b**, Stereo view of the ribosomal tunnel exit (white star). A model of the nascent chain with TMH anchor is shown in red. **c**, Contact 3. Nucleotides 72–76 of 4.5S RNA and L18 are shown in green, and the crosslink site U84 and nucleotides

2,828–2,837 of 23S RNA are shown in red²⁵. **d**, Contact 4 and conformational changes in helix 59. 23S RNA as observed in the crystal structure¹⁴ is shown in grey, helix 59 conformation in the RNC–SRP complex in green, and the Ffh M domain in yellow. The density is displayed at high contour level ($\sigma = 2.5$).

the 25% reduction in affinity of SRP for a TF–ribosome complex¹³, and the decrease in crosslinking efficiency between SRP position 17 and L23 in the presence of TF¹².

The second SRP–ribosome contact (c2) is mediated mainly by helix 24 of 23S RNA and to a lesser extent by proteins L24 and L22. On the SRP side, the contact is formed by the C-terminal helix M5 of the M domain and could also involve the linker between helices M3 and M4. Helix M5 is probably also involved in the binding of the signal anchor⁸.

The third contact site (c3) is furthest from the polypeptide tunnel exit. It involves ribosomal protein L18 and 4.5S RNA, probably nucleotides 72–76. These residues are part of domain IV and of the kink region of 4.5S RNA (Fig. 2a). C3 is adjacent to nucleotides 2,828–2,837 of 23S RNA, to which U84 of 4.5S RNA has been cross-linked in the presence of a nascent chain²⁵ (Fig. 3c).

In the fourth contact site (c4), we observed a pronounced (9 Å) movement of helix 59 of 23S RNA towards the SRP M domain (Fig. 3d). This rather large conformational change indicates that the ribosome is not a rigid platform for SRP binding. However, it is not clear whether it might serve to communicate the state of the ribosome to other factors.

The M domain of SRP has a hydrophobic groove, which was suggested to be part of the signal anchor-binding site of the nascent chain^{8,17}. Our structure shows that this groove is very close to the exit of the ribosomal tunnel (Fig. 3b). According to this model, the TMH is buried between the M domain and the ribosome. Conversely, in the absence of the signal sequence, the SRP may only contact the ribosome at protein L23 via its N domain, which has been characterized as the ribosome-binding domain of SRP^{6,22} (Fig. 4). This is in agreement with our reconstruction of a 70S–SRP complex without a

nascent chain, in which SRP density was only present at c1 (see Supplementary Fig. 3). This interaction would allow the M domain to scan the nascent peptide for a signal sequence in a translating ribosome, which results in full SRP docking at c2–c4 upon recognition of a signal sequence, as visualized in the SRP–RNC reconstruction.

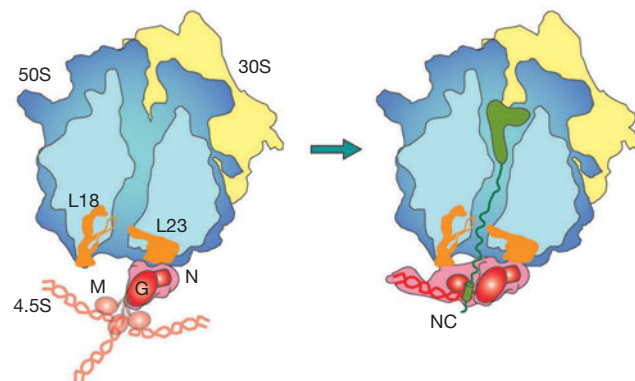


Figure 4 | Model of signal anchor-dependent docking of SRP to the ribosome. In the absence of a nascent chain, the SRP binds via its N domain to L23, and the M domain scans for a nascent chain with a signal anchor. In this state, the M domain with Ffs is flexible. Upon recognition of a signal anchor, the SRP binds tightly to the ribosome and forms three additional contacts with 50S. The signal anchor is buried in the hydrophobic pocket of the M domain, and the NG domain has an increased affinity for GTP, which is a prerequisite for FtsY binding^{2–30}. 30S, yellow; 50S, blue; peptidyl-tRNA, green; SRP, red; contact proteins, orange.

This is further supported by the RNA–RNA crosslink next to c3 (Fig. 3c), which is seen only in the presence of a nascent chain with a signal sequence²⁵. The formation of additional contacts coupled with the interactions between the hydrophobic signal sequence and the M domain of Ffh would explain salt-resistant binding of the SRP to the translating ribosome with picomolar affinity (Fig. 4).

We have previously determined the cryo-EM structure of the *E. coli* RNC–translocon complex at 15 Å resolution¹⁰. Together, these two structures set the stage for further structural studies aimed at deciphering at the molecular level the entire process of co-translational targeting and translocation in prokaryotes.

METHODS

SRP–RNC complexes were generated and characterized as described in the Supplementary Information. For grid preparation, RNCs (100 nM final concentration) were incubated on ice for 30 min with a 19-fold molar excess of SRP in 50 mM HEPES–KOH pH 7.5, 100 mM KCl, 25 mM MgCl₂, 1 mM DTT and 1 mM GTP. Micrographs were recorded at a magnification of ×50,000 under low-dose conditions with an FEI Tecnai F20 electron microscope operated at 200 kV. Micrographs were scanned on a Nikon super coolscan 9000 scanner, corresponding to a pixel size of 1.27 Å on the object scale. Subsequently, the data were 3× binned to 3.81 Å per pixel. 63,000 RNC particles were picked manually using X3d, CTF-corrected by CTF-Mix²⁶, and analysed with the SPIDER software package²⁷. Supervised classification²⁸ computationally separated empty ribosomes from SRP-bound ribosomes. 27,400 particles (43.5%) had the same Euler angles for the alignment against the empty ribosome and the SRP-bound ribosome and a better correlation coefficient for the SRP-bound ribosome (see also Supplementary Information). The final resolution was assessed by Fourier shell correlation characteristics: ~11 Å at 3σ and ~16 Å at FSC 0.5 (See Supplementary Fig. 1 and Supplementary Table 1). The map was low-pass filtered to 11 Å (the 3σ resolution). The X-ray structure of 70S (main body, L7/12 and L1 stalk)¹⁴ was fitted as rigid bodies using the program SITUS version 2.2¹⁵. Helix 59 was readjusted manually using O²⁹. The *T. aquaticus* Ffh NG domain¹⁶, the *S. solfataricus* M domain with RNA¹⁷, and the remainder of the 4.5S RNA¹⁸ were placed manually using O²⁹. The long flexible linker of *S. solfataricus*¹⁷ was rebuilt using O. The fit of SRP was further refined using NOMAD for normal mode flexible fitting¹⁹ (see Supplementary Information) and was energy-minimized using CNS Version 1.1²⁰.

Received 22 May; accepted 18 August 2006.

Published online 29 October 2006.

1. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775 (2001).
2. Nagai, K. *et al.* Structure, function and evolution of the signal recognition particle. *EMBO J.* **22**, 3479–3485 (2003).
3. Doudna, J. A. & Batey, R. T. Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* **73**, 539–557 (2004).
4. Ulbrandt, N. D., Newitt, J. A. & Bernstein, H. D. The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* **88**, 187–196 (1997).
5. Walter, P. & Blobel, G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**, 691–698 (1982).
6. Gu, S. Q., Peske, F., Wieden, H. J., Rodnina, M. V. & Wintermeyer, W. The signal recognition particle binds to protein L23 at the peptide exit of the *Escherichia coli* ribosome. *RNA* **9**, 566–573 (2003).
7. Pool, M. R., Stumm, J., Fulga, T. A., Sinning, I. & Dobberstein, B. Distinct modes of signal recognition particle interaction with the ribosome. *Science* **297**, 1345–1348 (2002).
8. Keenan, R. J., Freymann, D. M., Walter, P. & Stroud, R. M. Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* **94**, 181–191 (1998).
9. Halic, M. *et al.* Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* **427**, 808–814 (2004).
10. Mitra, K. *et al.* Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* **438**, 318–324 (2005).

11. Valent, Q. A. *et al.* Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. *Mol. Microbiol.* **25**, 53–64 (1997).
12. Buskiewicz, I. *et al.* Trigger factor binds to ribosome–signal-recognition particle (SRP) complexes and is excluded by binding of the SRP receptor. *Proc. Natl Acad. Sci. USA* **101**, 7902–7906 (2004).
13. Raine, A., Ivanova, N., Wikberg, J. E. & Ehrenberg, M. Simultaneous binding of trigger factor and signal recognition particle to the *E. coli* ribosome. *Biochimie* **86**, 495–500 (2004).
14. Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
15. Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).
16. Padmanabhan, S. & Freymann, D. M. The conformation of bound GMPNP suggests a mechanism for gating the active site of the SRP GTPase. *Structure* **9**, 859–867 (2001).
17. Rosendal, K. R., Wild, K., Montoya, G. & Sinning, I. Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc. Natl Acad. Sci. USA* **100**, 14701–14706 (2003).
18. Rosenblad, M. A., Gorodkin, J., Knudsen, B., Zwieb, C. & Samuelsson, T. SRPDB: Signal Recognition Particle Database. *Nucleic Acids Res.* **31**, 363–364 (2003).
19. Lindahl, E., Azuara, C., Koehl, P. & Delarue, M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* **34**, W52–W56 (2006).
20. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921 (1998).
21. Wild, K., Halic, M., Sinning, I. & Beckmann, R. SRP meets the ribosome. *Nature Struct. Mol. Biol.* **11**, 1049–1053 (2004).
22. Ullers, R. S. *et al.* Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the *Escherichia coli* ribosome. *J. Cell Biol.* **161**, 679–684 (2003).
23. Eisner, G., Moser, M., Schäfer, U., Beck, K. & Müller, M. Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. *J. Biol. Chem.* **281**, 7172–7179 (2006).
24. Ferbitz, L. *et al.* Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).
25. Rinke-Appel, J. *et al.* Crosslinking of 4.5S RNA to the *Escherichia coli* ribosome in the presence or absence of the protein Ffh. *RNA* **8**, 612–625 (2002).
26. Conway, J. F. & Steven, A. C. Methods for reconstructing density maps of “single” particles from cryoelectron micrographs to subnanometer resolution. *J. Struct. Biol.* **128**, 106–118 (1999).
27. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
28. Valle, M. *et al.* Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J.* **21**, 3557–3567 (2002).
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A 47**, 110–119 (1991).
30. Bacher, G., Lütcke, H., Jungnickel, B., Rapoport, T. A. & Dobberstein, B. Regulation by the ribosome of the GTPase of the signal-recognition particle during protein targeting. *Nature* **381**, 248–251 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank all members of the Ban laboratory for discussions and help with programs, and C. Frick, L. Verschragen (deceased) and J. Onderwater for technical assistance. W. Wintermeyer provided pET24a_Ffh. T. Shaikh and J. Frank are acknowledged for a script for supervised classification. We thank the Electron Microscopy Center Zürich (EMEZ) for support. C.S. was supported by a postdoctoral fellowship from the Ernst Schering Research Foundation. R.I.K. was supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO). This work was supported by the Swiss National Science Foundation (SNSF), the NCCR Structural Biology program of the SNSF, and a Young Investigator grant from the Human Frontier Science Program (to N.B.).

Author Information Coordinates of the atomic model of SRP have been deposited in the Protein Data Bank under the accession code 2iy3. The cryo-EM maps have been deposited in the 3D-EM database (EMBL—European Bioinformatics Institute, Cambridge, UK) under accession numbers EMD-1250 and EMD-1251. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for material should be addressed to N.B. (ban@mol.biol.ethz.ch).

Following the signal sequence from ribosomal tunnel exit to signal recognition particle

Mario Halic^{1*}, Michael Blau^{1*}, Thomas Becker¹, Thorsten Mielke², Martin R. Pool³, Klemens Wild⁴, Irmgard Sinning⁴ & Roland Beckmann^{1,2}

Membrane and secretory proteins can be co-translationally inserted into or translocated across the membrane¹. This process is dependent on signal sequence recognition on the ribosome by the signal recognition particle (SRP), which results in targeting of the ribosome–nascent-chain complex to the protein-conducting channel at the membrane^{2,3}. Here we present an ensemble of structures at subnanometre resolution, revealing the signal sequence both at the ribosomal tunnel exit and in the bacterial and eukaryotic ribosome–SRP complexes. Molecular details of signal sequence interaction in both prokaryotic and eukaryotic complexes were obtained by fitting high-resolution molecular models. The signal sequence is presented at the ribosomal tunnel exit in an exposed position ready for accommodation in the hydrophobic groove of the rearranged SRP54 M domain. Upon ribosome binding, the SRP54 NG domain also undergoes a conformational rearrangement, priming it for the subsequent docking reaction with the NG domain of the SRP receptor. These findings provide the structural basis for improving our understanding of the early steps of co-translational protein sorting.

Elongation-arrested *Escherichia coli* 70S ribosomes were purified from an *in vitro* translation system⁴. As a nascent chain we used the first 102 amino acid residues of the membrane protein FtsQ, which contain a single transmembrane segment serving as an SRP-dependent signal sequence⁵. Purified 70S ribosome–nascent-chain complexes (RNCs) were reconstituted with recombinant SRP and subjected to structure determination by cryo-electron microscopy and single-particle analysis. To reach subnanometre resolution, sorting of the data sets⁶ was necessary.

The structure of a 70S RNC carrying a signal sequence was reconstructed at 9.1 Å resolution from particles that were combined after sorting on the basis of the presence of P-site tRNA and the absence of a ligand. The reconstruction showed the expected P-site peptidyl-tRNA and, in addition, an elongated density located directly at the ribosomal tunnel exit site (Fig. 1a, b). This density is not present in comparable reconstructions at 9 Å resolution⁷ such as that of a stalled *E. coli* 70S ribosome without a nascent chain (Fig. 1c). Because extended peptide chains would hardly be resolved at the present resolution, this density most probably represents the signal sequence of FtsQ in an α -helical conformation, which it might already have adopted in the ribosomal tunnel^{8,9}. The density is in contact with a protruding hairpin structure of ribosomal protein L24p. It is also in the immediate vicinity of the rRNA helices 59 and 24, and the proteins L23p and L29p (Fig. 1d), which is in excellent agreement with chemical crosslinks between FtsQ and L23p (ref. 5). Because the density did not show the clear rod-like shape as observed for α -helices in several ribosomal proteins of this reconstruction, it is likely that

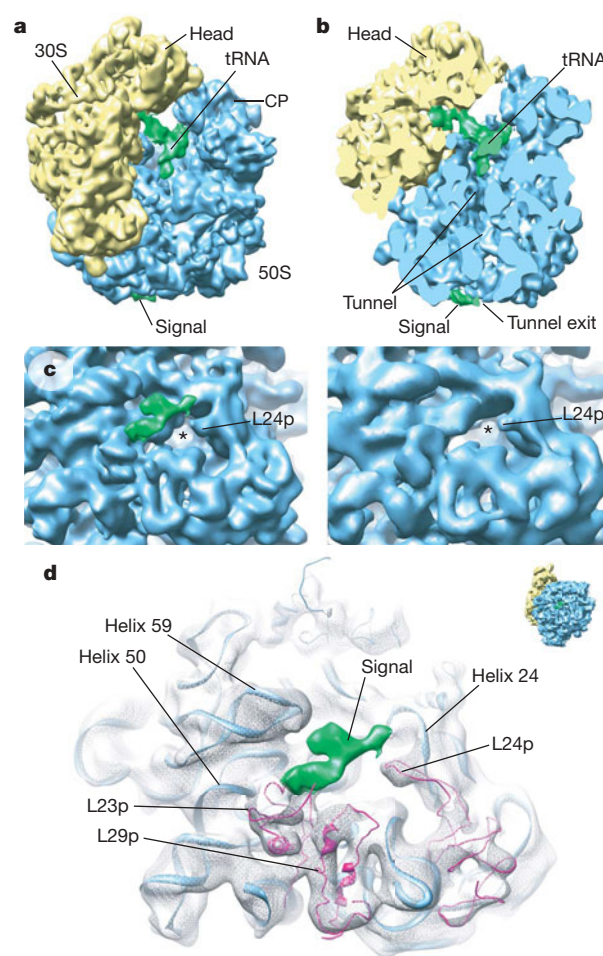


Figure 1 | Cryo-electron microscopic structure of programmed 70S ribosome (RNC) with signal sequence. **a**, Density map with the ribosomal 30S subunit shown in yellow, 50S in blue and the P-site tRNA and nascent chain (signal sequence) in green. Landmarks are indicated. CP, central protuberance. **b**, Section through the ribosome showing the signal sequence at the tunnel exit site. **c**, Comparison of the ribosomal tunnel exit site of 70S RNC (left) with the empty 70S ribosome (3D-EM database accession number EMD1055)⁷ (right). Additional density (signal sequence) is shown in green. **d**, View as in **c**, with ribosomal density shown as a white mesh and models as ribbons. rRNA is shown in blue; L23p, L24p and L29p in magenta; signal sequence in green.

¹Gene Center, Department of Chemistry and Biochemistry, University of Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. ²UltraStructureNetwork, USN, Max Planck Institute for Molecular Genetics, Ihnestrasse 63–73, 14195 Berlin, Germany. ³Faculty of Life Sciences, Michael Smith Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK. ⁴Heidelberg University Biochemistry Center (BZH), Im Neuenheimer Feld 328, D-69120 Heidelberg, Germany.

*These authors contributed equally to this work.

the signal sequence confers some flexibility at this site. Targeting factors or chaperones can easily interact with the signal sequence in its exposed position. Because SRP can interact with signal sequences in nascent chains of various lengths^{10,11}, the observed site could serve as a parking position at the tunnel exit where signal sequences are retained for recognition by SRP¹².

The reconstruction of the *E. coli* 70S RNC–SRP complex at 9.4 Å revealed a large density at the tunnel exit representing SRP (Fig. 2a–c). As a result of conformational heterogeneity and imperfect occupancy, the local resolution of the SRP is somewhat lower than that of the ribosome (9.5–12 Å). The elongated portion showed a helical twist, with the major and minor groove representing 4.5S RNA of SRP. Attached to it is the SRP54 (Ffh in *E. coli*) by means of its well-

resolved M domain¹³ and, barely visible at subnanometre resolution, the SRP54 NG domain. It seems to have some freedom to move and is clearly visible only after the application of a low-pass filter, reducing the information to about 13 Å (Fig. 2b). We docked available high-resolution X-ray data^{13,14} into the density and flexibly adjusted minor parts to build a molecular model of ribosome-bound *E. coli* SRP. Although the RNA of the X-ray data is truncated at loop C, we can follow it until loop E. Notably, the visible part of the 4.5S RNA has a kink of about 30° formed by the helix around nucleotide 72 next to loop C (Fig. 2c, f). Some flexibility may be provided by loop C in the RNA, and total loss of density after loop E indicates that the latter serves as a very flexible hinge¹⁵. Although *E. coli* SRP is positioned a few ångströms closer to the ribosome than is the mammalian SRP,

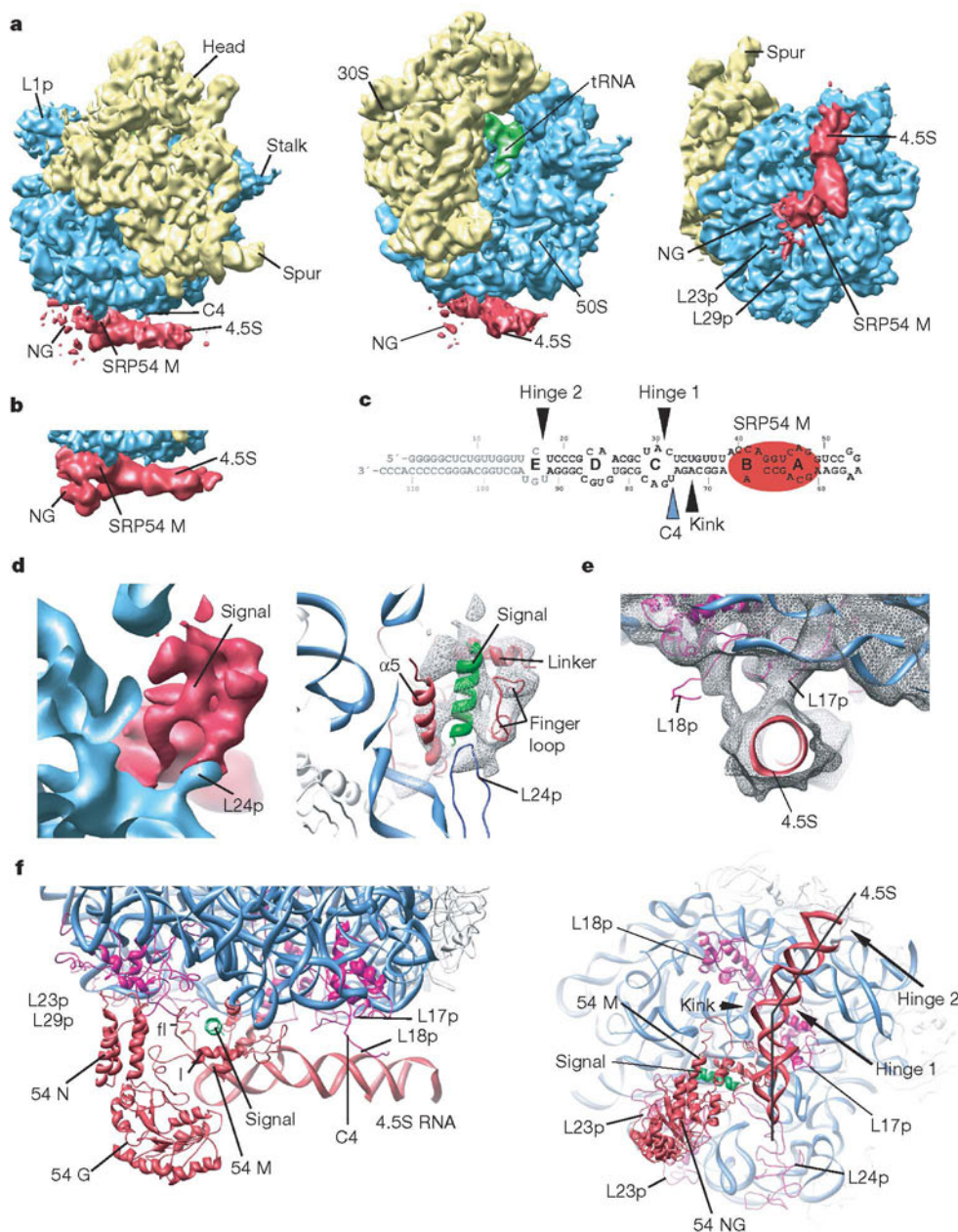


Figure 2 | Structure of *E. coli* 70S RNC–SRP complex. **a**, Density map with colour code as in Fig. 1 and SRP shown in red, left, view onto the back of the 30S subunit; middle, side view; right, bottom view. Landmarks are indicated. **b**, SRP density filtered at a lower resolution (13 Å) and shown at a lower contour level. Note the visible NG domain. **c**, Secondary structure of *E. coli* 4.5S RNA. Red, binding region of the SRP54 M domain; grey, flexible part of

SRP RNA not visible in the cryo-electron microscopic density. Arrowheads indicate hinges, kinks and interaction sites. **d**, Hydrophobic groove of the SRP54 M domain with signal sequence bound (left), and the model of SRP54 M domain (red) and the signal sequence (green) docked into the density (right). rRNA is shown in blue, L24p in dark blue. **e**, Connection 4 of *E. coli* SRP. **f**, Left, side view of the SRP–RNC model; right, bottom view.

the connections of *E. coli* SRP with the 70S ribosome are very similar to those found for the mammalian S domain^{4,16}. The SRP54 NG domain interacts through the N domain with the L23p/L29p adaptor site, and the SRP54 M domain interacts with ribosomal RNA helices 24, 59 and 50. As in the eukaryotic system, the M domain might also interact with L22p (rpL17 in yeast). An entirely new contact is established between L24p and the finger loop of the SRP54 M domain (Fig. 2d), resembling the L24p contact between ribosome and TF¹⁷. In a similar manner to connection 4 of mammalian SRP, the SRP RNA forms a contact around nucleotide 75 in loop C of the 4.5S RNA with positively charged residues of the ribosomal protein L18p and probably also with the carboxy terminus of L17p (Fig. 2b, e). We also observed density for the linker helix connecting the M domain with the NG domain of SRP54. It is in contact with α -helix 1 and the finger loop of the SRP54 M domain as well as with the NG domain, thus efficiently coupling the signal-sequence-binding domain with the GTPase domain of SRP54 as described¹⁸.

Notably, additional density is present in the hydrophobic groove of the SRP54 M domain (Fig. 2d), which has previously been suggested as the main signal-sequence-binding site of SRP¹⁹. This density, thus very probably representing the signal sequence, was

observed in a position that in unbound SRP is occupied by a part of the finger loop of the SRP54 M domain. This may serve as a pseudo-substrate in the absence of a signal sequence. Because we observed only weak density for the rearranged finger loop in the RNC–SRP complex, we concluded that it does not rearrange into a distinct conformation after being pushed out of the hydrophobic groove. Although the density for the signal sequence is not reaching down to the backbone of the SRP RNA, because of possible flexibility we cannot exclude an interaction with RNA¹³.

The position of the signal sequence in the RNC–SRP complex is very similar to that of the ligand-free RNC-bound signal sequence (Fig. 3). Only its contact with L23p was lost in the RNC–SRP reconstruction, explaining the loss of crosslinks between L23p and the FtsQ signal sequence after SRP binding⁵. The interaction of SRP with the ribosomal adaptor site L23p/L29p may trigger the dissociation of the signal sequence and, concomitantly, the interaction of L24p with the finger loop of the SRP54 M domain may contribute to the opening of the hydrophobic groove so that the signal sequence can glide into it (Fig. 2d). However, for switching from the ribosome-bound state to the SRP-bound state the signal sequence has to move very little: SRP54 is merely locking it in by providing a matching hydrophobic environment directly at its exposed position at the tunnel exit.

We improved the cryo-electron microscopic structure of the 80S RNC in complex with the mammalian SRP⁴ to a resolution of 8.7 Å (Fig. 4a). Although the local resolution of the SRP S domain is estimated to be in the range 9.5–11.5 Å, the density permitted the fitting of a molecular model for SRP54 based on α -helical secondary structure. In a very similar manner to the 70S RNC–SRP complex, docking of the M domain of mammalian SRP54 revealed additional density in the hydrophobic groove, representing the bound signal sequence (Fig. 4b, c). In comparison with the M domain in unbound SRP, the engaged M domain is rearranged when accommodating the signal sequence: the α -helix 1 is rotated upwards while still keeping the contact to the repositioned linker helix, which might involve the ‘greasy slide’ proposed earlier¹⁴. In this way the groove is formed by α -helices 1, 2 and 5, the finger loop and the linker helix. Contacts were also established between the linker helix and the finger loop, and between the G domain and α -helix 2 of the M domain. These contacts may contribute to the communication between the M and G domains of SRP54.

In comparison with the bacterial SRP54 subunit, the mammalian protein has a C-terminal extension of the M domain of about 70 amino acids, which is predicted to adopt a primarily α -helical secondary structure. Using protein threading we created a molecular model of this C terminus that could be fitted into the remaining density with only very minor modifications (Fig. 4d). It wraps around the ribosomal rRNA helix 59 and reaches up towards the tunnel exit, also interacting with ribosomal rRNA helix 50. In this position the C terminus confines the hydrophobic groove on one side and is very likely to contribute to the interaction of the SRP54 M domain with the signal sequence.

The overall geometry of the refitted SRP54 N domain¹⁴, in particular α -helices 2 and 4, properly filled the density, but with α -helices 1 and 3 partly shifted outside (Fig. 4e, f). When using the N domain of SRP54 NG involved in the twin structure with the NG domain of the SRP receptor^{20,21}, it became evident that here α -helix 3 fits better, but α -helix 1 is delocalized and α -helix 4 no longer fits as a result of a rotation. This indicates that the N domain of SRP undergoes a conformational switch upon ribosome interaction, which leads to an arrangement of α -helices 2 and 3 very similar to the conformation observed in the NG twin structure. These two helices and their connecting loop include the highly conserved ‘ALLEADV’ motif that is involved in the interface between the two NG domains in the NG twin^{20,21}. Thus, high-affinity ribosome binding seems to prime this part of the NG domain for the subsequent interaction with the SRP receptor. The reorientation of α -helices 1 and 4 of the SRP54 N domain after NG twin formation may explain how the interaction

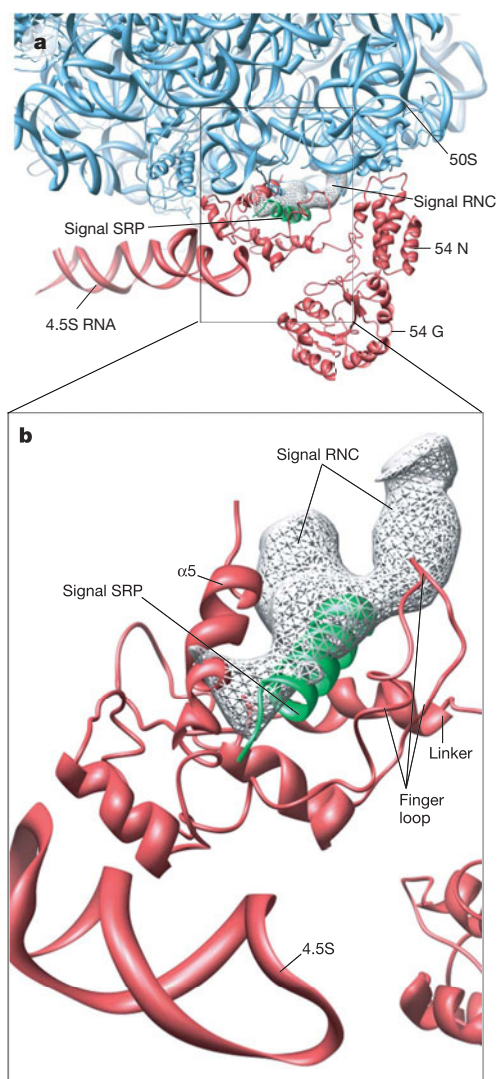


Figure 3 | Transfer of signal sequence from the ribosome to SRP. **a**, The models of the ribosome, SRP and the SRP-bound signal sequence are shown in blue, red and green, respectively. The density of the signal sequence bound to the ribosome (Fig. 1) is shown as a white mesh. **b**, Close-up of **a**. Note the small difference and partial overlap of the positions of the signal sequence.

with the SRP receptor triggers dissociation of the SRP54 NG domain from its ribosomal binding site²².

Moreover, when adjusting the angle between the N and G domains of SRP54 we observed that the angle is very similar to that observed in the NG twin structure (Fig. 4g). Rearrangement from the monomeric to the NG twin conformation requires a rotation of about 30°; a rotation of about 25° already takes place when switching to the ribosome-bound conformation, which may also contribute to the regulation of the affinity for GTP. The entire conformation of the SRP54 NG domain is therefore rearranged on the ribosome such that

it is primed for the interaction with the NG domain of the SRP receptor.

Although the resolution of the NG domain is not as good, these observations are consistent with the structure of the bacterial system. When the 70S RNC-bound SRP and the 80S RNC-bound SRP are superimposed (Fig. 4h, i), it is evident that these systems are extremely similar in their overall conformation. This indicates a high degree of structural and functional conservation, which is underlined by the fact that *in vitro* the mammalian targeting system can be functionally replaced by the bacterial one²³.

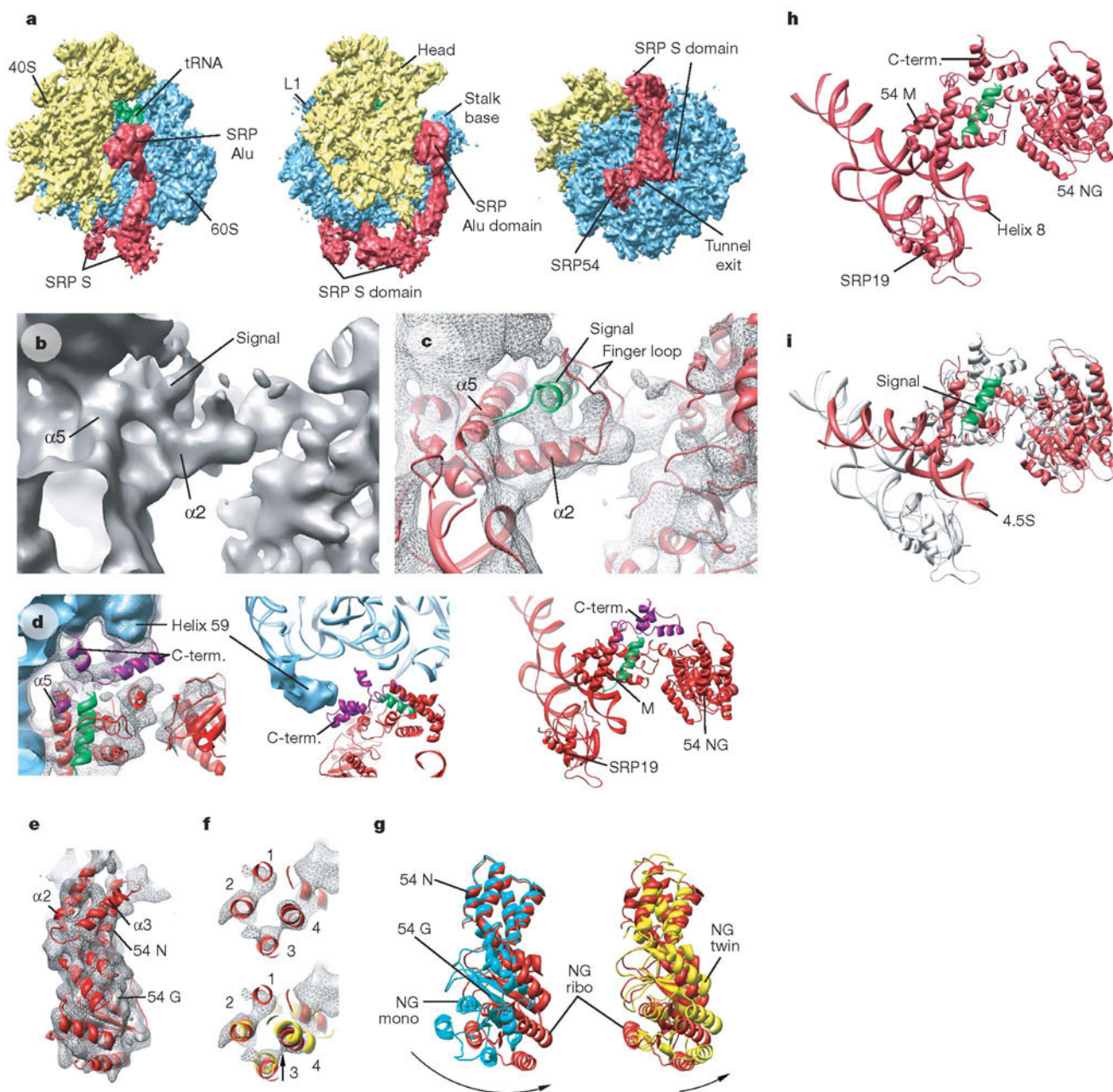


Figure 4 | Mammalian SRP bound to 80S RNC. **a**, Cryo-electron microscopic structure of the mammalian SRP bound to 80S wheat germ RNC. **b**, Density representing the hydrophobic groove of the SRP54 M domain with the signal sequence bound. **c**, Same view as in **b** showing the docked crystal structure of SRP54 M domain (red) and the signal sequence (green) docked into extra density. **d**, A model of the C-terminal part of SRP54 M domain is shown in purple docked into the additional density of SRP54 (left). SRP54 density is shown as a grey mesh, and the ribosome is shown in blue. **e**, Model of the fitted ribosome-bound mammalian NG

domain. **f**, Top view of the four-helix bundle of the SRP54 N domain based on the monomeric NG domain in red (top), and superimposed on the N domain from the NG twin structure in yellow (bottom). **g**, Comparison of the model of ribosome-bound NG domain (red, NG ribo) with the monomeric unbound domain (blue, NG mono) and with the twin NG domain (yellow, NG twin). **h**, The complete model of mammalian SRP S domain (red). **i**, The *E. coli* model (red) superimposed on the mammalian model (white).

METHODS

E. coli 70S RNCs were purified and reconstituted essentially as described previously⁴, with some modifications (see Supplementary Information). Samples were applied to carbon-coated holey grids²⁴ and micrographs were recorded on a Tecnai F30 microscope at 300 kV. The data were processed with SPIDER²⁵ and classified⁶. As a result, data sets of stalled ribosomes with a peptidyl-tRNA in the P-site were further split in a manner that was dependent on the presence of ligands or distinguishable conformations. The crystal structure of *E. coli* 50S subunit (Protein Data Bank (PDB) accession number 2AW4)²⁶ was docked with the use of Situs²⁷. The *E. coli* SRP (1DUL)¹³, *S. sulfataricus* SRP (1QZW)¹⁴ and mammalian SRP structure were docked with the use of O²⁸, and the flexible amino-terminal part of the M domain was adjusted accordingly. Figures were prepared with Chimera²⁹.

Received 24 July; accepted 10 October 2006.

Published online 29 October 2006.

- Luirink, J., von Heijne, G., Houben, E. & de Gier, J. W. Biogenesis of inner membrane proteins in *Escherichia coli*. *Annu. Rev. Microbiol.* **59**, 329–355 (2005).
- Luirink, J. & Sinning, I. SRP-mediated protein targeting: structure and function revisited. *Biochim. Biophys. Acta* **1694**, 17–35 (2004).
- Halic, M. & Beckmann, R. The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.* **15**, 116–125 (2005).
- Halic, M. *et al.* Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* **427**, 808–814 (2004).
- Ullers, R. S. *et al.* Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the *Escherichia coli* ribosome. *J. Cell Biol.* **161**, 679–684 (2003).
- Penczek, P. A., Frank, J. & Spahn, C. M. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol.* **154**, 184–194 (2006).
- Valle, M. *et al.* Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nature Struct. Biol.* **10**, 899–906 (2003).
- Woolhead, C. A., McCormick, P. J. & Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725–736 (2004).
- Lu, J. & Deutsch, C. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **44**, 8230–8243 (2005).
- Houben, E. N., Zarivach, R., Oudega, B. & Luirink, J. Early encounters of a nascent membrane protein: specificity and timing of contacts inside and outside the ribosome. *J. Cell Biol.* **170**, 27–35 (2005).
- Flanagan, J. J. *et al.* Signal recognition particle binds to ribosome-bound signal sequences with fluorescence-detected subnanomolar affinity that does not diminish as the nascent chain lengthens. *J. Biol. Chem.* **278**, 18628–18637 (2003).
- Eisner, G., Moser, M., Schafer, U., Beck, K. & Muller, M. Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. *J. Biol. Chem.* **281**, 7172–7179 (2006).
- Batey, R. T., Rambo, R. P., Lucast, L., Rha, B. & Doudna, J. A. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**, 1232–1239 (2000).
- Rosendal, K. R., Wild, K., Montoya, G. & Sinning, I. Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc. Natl Acad. Sci. USA* **100**, 14701–14706 (2003).
- Buskiewicz, I. *et al.* Conformations of the signal recognition particle protein Ffh from *Escherichia coli* as determined by FRET. *J. Mol. Biol.* **351**, 417–430 (2005).
- Pool, M. R., Stumm, J., Fulga, T. A., Sinning, I. & Dobberstein, B. Distinct modes of signal recognition particle interaction with the ribosome. *Science* **297**, 1345–1348 (2002).
- Schlunzen, F. *et al.* The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure* **13**, 1685–1694 (2005).
- Wild, K., Halic, M., Sinning, I. & Beckmann, R. SRP meets the ribosome. *Nature Struct. Mol. Biol.* **11**, 1049–1053 (2004).
- Bernstein, H. D. *et al.* Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature* **340**, 482–486 (1989).
- Egea, P. F. *et al.* Substrate twinning activates the signal recognition particle and its receptor. *Nature* **427**, 215–221 (2004).
- Focia, P. J., Shepotinovskaya, I. V., Seidler, J. A. & Freymann, D. M. Heterodimeric GTPase core of the SRP targeting complex. *Science* **303**, 373–377 (2004).
- Halic, M. *et al.* Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science* **312**, 745–747 (2006).
- Powers, T. & Walter, P. Co-translational protein targeting catalyzed by the *Escherichia coli* signal recognition particle and its receptor. *EMBO J.* **16**, 4880–4886 (1997).
- Wagenknecht, T., Grassucci, R. & Frank, J. Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from *Escherichia coli*. *J. Mol. Biol.* **199**, 137–147 (1988).
- Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
- Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).
- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **A47**, 110–119 (1991).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the VolkswagenStiftung and the Deutsche Forschungsgemeinschaft SFB594 (to R.B.) and SFB638 (to I.S.) and by the European Union and Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin (UltraStructureNetwork).

Author Information Coordinates of the atomic models of SRP have been deposited in the PDB under accession numbers 2j28 and 2j37. The cryo-electron microscopic maps have been deposited in the 3D-EM database under accession numbers EMD1261 (*E. coli* SRP-RNC), EMD1263 (*E. coli* RNC) and EMD1264 (mammalian SRP-RNC). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B. (beckmann@lmb.uni-muenchen.de).

ADDENDUM

doi:10.1038/nature05366

Human embryonic stem cell lines derived from single blastomeres

Irina Klimanskaya, Young Chung, Sandy Becker, Shi-Jiang Lu & Robert Lanza

Nature 444, 481–485 (2006); doi:10.1038/nature05142 (this issue)

At the request of the Editors at *Nature*, we wish to clarify some questions that have arisen since the advance online publication (AOP) of our Letter on 23 August 2006. In our Letter, we showed that human embryonic stem-cell lines can be generated from a single cell after its removal from an 8–10-cell embryo. To minimize the number of embryos used, we removed multiple cells from each embryo, and none of the biopsied embryos were allowed to develop in culture.

In our experiments, the isolated blastomeres from each embryo were cultured together in the same medium that was used to culture the parent embryo, and were arranged to avoid contact with each other. Diffusible factors from the other blastomeres present in the media may assist recovery and growth of the blastomere. We have not excluded the possibility that only a subset of blastomeres of an 8–10-cell embryo are capable of forming human embryonic stem cells. These caveats are worth considering for future studies, but do not negate our central finding that blastomeres extracted from an 8–10-cell embryo by mechanical micromanipulation can form human embryonic stem-cell cultures.

We have now added more explicit information on how individual embryos were handled in the form of a table based on Supplementary Table 1 of the AOP version of the Supplementary Information (which has now been removed). This information is now presented in the printed paper as Table 1, to indicate how many cells were individually biopsied from each embryo. In addition, the descriptions for Fig. 4b and d in the legend to Fig. 4 have been corrected (they were inadvertently transposed in the AOP version of the paper).

These clarifications have been incorporated into the paper for the print version and are individually listed as Supplementary Information to this Addendum.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

CORRIGENDUM

doi:10.1038/nature05400

Insights into social insects from the genome of the honeybee *Apis mellifera*

The Honeybee Genome Sequencing Consortium

Nature 443, 931–949 (2006)

In this Article, the surname of co-author L. Sian Gramates was misspelled Grametes.

CORRIGENDUM

doi:10.1038/nature05423

Potential of stem-cell-based therapies for heart disease

Deepak Srivastava & Kathryn N. Ivey

Nature 441, 1097–1099 (2006)

It has been drawn to our attention (by J. Lakota, and by J. J. Minguell and G. P. Lasala) that we used the abbreviation for bone-marrow-derived stem cells (BMSCs) inappropriately in some parts of our Insight Progress article. BMSC is a commonly used acronym for the heterogeneous adult stem cells in bone marrow. The term BMSC was used with this intention, but was placed after describing a study on bone-marrow-derived mesenchymal stem cells, which are a subset of BMSCs. Subsequent references to BMSCs were intended to describe the heterogeneous cells, rather than the specific mesenchymal subtype. Most clinical trials for myocardial infarction used BMSCs and have so far had mixed results. Future trials with isolated mesenchymal stem cells will reveal their potential in the context of heart disease.

ERRATUM

doi:10.1038/nature05377

Eastern Pacific cooling and Atlantic overturning circulation during the last deglaciation

Markus Kienast, Stephanie S. Kienast, Stephen E. Calvert, Timothy I. Eglinton, Gesine Mollenhauer, Roger François & Alan C. Mix

Nature 443, 846–849 (2006)

In Figure 1 of this Letter, the units of organic carbon burial flux on the left y axis should be $\text{g m}^{-2} \text{yr}^{-1}$ and not $\text{g m}^{-2} \text{kyr}^{-2}$. In addition, an earlier version of the Supplementary Information for this Letter was inadvertently uploaded. The Supplementary Information was corrected on 30 October 2006.

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

German universities last month received a significant cash injection to help them improve funding for young investigators, build university-industry research hubs and attract scientists from outside the country. The German government's Excellence Initiative awarded €873 million (US\$1.1 billion) over five years, the first step in a €1.9-billion bid to strengthen science in Germany. The money is being divided between graduate schools, research clusters and key research programmes. The initiative was launched after policy-makers noticed how few German universities were represented in the top 100 rankings of international universities and that the country had difficulty attracting researchers from abroad.

The approach differs from traditional funding in that it singles out a relatively small number of institutions for large grants, decided by an international panel outside the country, rather than distributing money equally to universities throughout Germany. It aims to promote specialization in some areas, rather than have many universities work on similar research agendas. Three institutions were the big winners of the first round, receiving funds in all three of the initiative's categories — the Karlsruhe University of Technology, the University of Munich and the Munich University of Technology.

This first round of funding sees roughly €175 million a year being split between initiatives at 22 universities; 18 graduate schools, 17 clusters of excellence and three key research programmes will share the money. The next round of applications for grants is already under way.

At the very least, this funding largesse should attract the attention of young German scientists weighing the benefits of training at home or abroad. If the programme receives continued funding and demonstrates some early research success, it should also meet its goal of attracting researchers from outside Germany. If that happens, expect other European countries to set up similar initiatives.

Paul Smaglik, *Naturejobs* editor

Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestranguez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



SMALL TALK

Nanobiotechnology is a growing field, but will it emulate the biotech boom? **Virginia Gewin** investigates.

Andrew Pelling likes to rock out at the nanoscale. Working with a musician colleague, he applied his knowledge of physical chemistry and an atomic force microscope to record sounds of a living yeast cell oscillating. "The Dark Side of the Cell" was an artistic offshoot of his PhD work in which he managed to convert the nanoscale cell movements he had co-discovered into sound. Now, having turned down offers from Germany, Japan and the United States, Pelling is pursuing his research on cell mechanics at the new London Centre for Nanotechnology at University College London.

Like most students attracted to the nanoscale realm of biotechnology, Pelling uses a range of skills to cover new ground. Nanobiotechnology develops tools to explore the molecular mechanisms behind biological processes. With technology reaching the actual scale of biological events, three areas — drug-delivery systems, diagnostics and therapeutics — have come to the fore.

Local and national governments in the United States, Germany, Britain and Japan are providing money for nanobiotechnology research and development initiatives. The US\$1.6-billion National Nanotechnology Initiative in the United States, Japan's US\$1.1-billion investment and some US\$4.5 billion earmarked for nanoscience in the European Union are creating the research and training infrastructure needed to boost the field. By some estimates, two million nano-savvy workers will be needed in the next decade. Training opportunities are abundant, but, although the outlook is good, the number of faculty research positions and industry jobs may not yet be sustainable.

Carving out a career in nanobiotechnology requires more than the cursory knowledge of multiple fields. Mastering at least two of the field's three core disciplines — molecular biology, chemistry and physics — is mandatory. Although PhD programmes are starting up that meet the needs of such interdisciplinary students, most current graduate applicants are still proficient in only one discipline.

Fortunately, many training programmes expose students to more than one area. At the Georgia Institute of Technology in Atlanta, for example, the nanocardiology fellow programme (one of three nanobiotech initiatives started at the institute in the past two years) requires students to work in nanotechnology development as well as in a cardiology lab. The biomedical-engineering school is a hotbed of nano-activity, boasting a new nanocancer centre as well as a nanomedicine centre run jointly with nearby Emory University (see 'Thinking small'). Other training programmes, such as the PhD programme at the Johns Hopkins Institute for NanoBioTechnology in Baltimore, Maryland, require students to be members of two advisers' research groups. Both programmes also include coursework in developing business plans, market assessment and securing funding (see *Nature* **435**, 124–125; 2005).

Big opportunities

Given the potential of nanosensors or molecular imaging tools to improve tumour detection, cancer is one of the hottest areas for research and graduate training opportunities. The US National Science Foundation and National Cancer Institute recently funded four centres for integrative graduate education and research traineeship programmes. The variety of university programmes indicates the dynamic nature of nanobiotechnology research: the University of New Mexico in Albuquerque specializes in integrative nanoscience and microsystems; New Jersey's Rutgers University focuses on nanopharmaceutical engineering and science; and the University of Washington in Seattle has a programme focused on building a workforce that can use nanoscale research platforms to understand and diagnose disease.

Student demand is also beginning to drive the development of undergraduate programmes in nanobiotechnology. Although Chad Mirkin, director of the International Institute for Nanotechnology



In tune with the times: Andrew Pelling managed an unlikely marriage of nanotechnology, music and yeast.

at Northwestern University in Evanston, Illinois, isn't convinced that undergraduate programmes are yet as meaningful as a solid foundation in the disciplines of physics, biology and chemistry, he can attest to the 'nano-effect'. Applications for graduate school at Northwestern went up dramatically after his institute formed in 2000. Increasing demand for training opportunities has also led to the formation of short courses such as the one on nanoelectronics at the European School on Nanosciences and Nanotechnologies in Grenoble, France. And the Georgia Institute of Technology offers a nanoscience and technology certificate programme to students of any undergraduate or graduate major. Only a few certificate-holders have so far graduated, but student participation is high — from a wide range of disciplines.

Given the early state of the field, skilled postdocs are in particularly high demand, says Jean-Marc Grognet, scientific director at both Nano2Life, a €15-million (US\$19-million) network of excellence, and NanoBio, a Grenoble-based initiative focused on developing rapid, nano-sized laboratories on a chip.

The dearth of highly trained people has encouraged the relatively few existing companies to fund PhD projects. Flemming Besenbacher, director of the Interdisciplinary Nanoscience Centre at the University of Aarhus in Denmark, says one-third of the PhD projects at the university are financed by industries involved with the centre. And global electronics powerhouse Philips is investing heavily in molecular imaging as one of the main private funders of the public-private partnership initiative forming the Center for Translational Molecular Medicine located in Eindhoven, the Netherlands.

Market watch

Despite the many training opportunities, the job market remains uncertain. "It's too early to say whether nanobiotech efforts will translate into the industrial equivalent of the first biotech boom," says Michael Horton, a nanobiotechnologist at the London Centre for Nanotechnology. More and bigger grants are allowing for the number of positions in some faculty departments to be increased — although some of these posts are taken by internal academics realigning themselves in the field.

One area where there is likely to be a shortfall is in the number of skilled clinicians who can conduct the testing needed for nanobio-based diagnostics or therapeutics. Grognet believes that doctors who can



Jean-Marc Grognet (top) and Chad Mirkin both see a high demand for skills in nanobiotechnology.

discuss nanobiotechnology with physicists or biologists will be needed to prepare the development of clinical trials for new products.

Businesses want students who can take the initiative — especially as formal nanobiotechnology positions at existing companies tend to be relatively limited.

Nanosphere, a nano life-sciences start-up in Northbrook, Illinois, is recruiting roughly 10 people to bring its total number of employees to 100. The firm's chief executive, Bill Moffitt, says there is ample talent available but notes that it is hard to find people willing to put in the long hours and assume the high risks that come with working for a smaller company.

Invitrogen, an established life-sciences company in Carlsbad, California, with thousands of employees, plans to add just ten nanobiotechnology specialists. In an effort to recruit the top European talent while avoiding the difficulties of getting US visas, Invitrogen is also opening a biotechnology manufacturing facility, with a nanobiotech component, in Glasgow, UK, in early 2007. The company's strategy includes funding for internship positions.

Mark Morrison, scientific manager of the UK Institute for Nanotechnology in Stirling, points out that few companies are working in nanobiotechnology in Europe, but he expects to see more in the near future. Indeed, Europe has only half as many companies as the United States, but there are more than 150 small companies worldwide pursuing nanobiotechnology/nanomedicine products.

Boom or bust?

But until nanobiotechnology products can demonstrate success in clinical trials and are subsequently embraced by the marketplace, the future of the field — and the estimates that it will create thousands of jobs — remains uncertain. The area is now at a stage akin to biotech companies before the completion of the draft human genome in 2001. Companies related to genomics then received disproportionate media attention and high stockmarket valuations. But when few genomics-related products materialized, the biotech stockmarket bubble burst.

As a result of this lesson from history, the larger pharmaceutical companies are hedging their bets — letting the smaller companies take the risks, while only dabbling in the field. GlaxoSmithKline, for example, is not currently recruiting for nanobiotechnology-specific talent. Nevertheless, Denny Van Liew, senior director of strategic management for Pfizer Global Research and Development, hints that Pfizer may get more active sooner rather than later — particularly as it forges collaborative relationships with a number of universities, notably the University of Michigan in Ann Arbor and the Massachusetts Institute of Technology in Cambridge.

The potential that nanobiotechnology has in drug delivery, diagnostics and therapeutics will continue to engage the interest, if not involvement, of the large companies. In the near term, global investment in nanobiotechnology should give large companies plenty of projects to follow. Young scientists working in the field can benefit from this scrutiny — once they demonstrate results that have potential applications. "At the end of the day, people are looking for creative minds," says Pelling. For now, at least, Pelling is in a class of his own — nanobiotechnologist by day, cell jockey by night.

Virginia Gewin is a freelance writer in Portland, Oregon.

THINKING SMALL

As part of its effort to speed scientific discoveries from the bench to the bedside, the US National Institutes of Health recently completed its nanomedicine initiative — a network of eight national nanomedicine-development centres. The Georgia Institute of Technology in Atlanta — along with Purdue University in Lafayette, Indiana; the University of California, Los Angeles; and the Lawrence Berkeley National Laboratory — join four centres funded last year. The initiative will fund research and training efforts using an engineering approach to improve understanding of nanoscale molecular complexes that should be key to the development of diagnosis and treatment strategies.

V.G.

► <http://nihroadmap.nih.gov/nanomedicine/index.asp>

MOVERS

Robert Huber, professor, Cardiff University, Cardiff, UK



1971–2005: Director, structural-research department, Max Planck Institute of Biochemistry, Martinsried, Germany
1976–present: Professor, Technische University, Munich, Germany
2006–present: Visiting professor, University of Duisburg-Essen, Germany

Robert Huber's path to a Nobel chemistry prize wasn't immediately clear, as his secondary school in Munich focused on Latin and Greek. But his early fascination with the chemistry of minerals, born on walks in the nearby mountains, soon asserted itself. Studying chemistry at the Technical University of Munich in the 1950s, he delved into the newly developing field of crystallography.

Huber's experimental drive was the perfect complement to the theoretical focus of his doctoral mentor, Walter Hoppe. Huber elaborated on Hoppe's proposed method for molecular crystallography by experimentally determining crystal structures using molecular fragments. His dissertation elucidating the insect metamorphosis hormone, ecdysone, boosted his reputation. Switching to protein analysis, he deciphered the crystal structure of the insect protein erythrocrucorin, a blood protein similar to mammalian proteins. Prestigious offers followed, and he began a 30-year career as director of structural research at the Max Planck Institute of Biochemistry in Martinsried.

Taking advantage of refined crystallography methods and instrumentation, Huber teamed up with his former PhD student Johann Deisenhofer and colleague Hartmut Michel to tackle one of the most important molecular structural targets in biology — the photosynthetic reaction centre, which converts solar rays into electrical energy. Determining its structure, the first one determined for a crystalline membrane protein, earned them the 1988 Nobel Prize in Chemistry. This didn't significantly affect Huber's career, but it did help him attract even better students.

After retiring at 68, as dictated by German law, Huber recently accepted a part-time position to help form an interdisciplinary chemical-biology initiative: a programme linking the schools of chemistry and biosciences at Cardiff University in Wales. He will teach classes as well as advise on strategic investments and recruitment.

"He's helped us find a focus — the structural basis for drug development — to build our activities around," says Adrian Harwood, a biosciences professor at Cardiff. He adds that Huber's excellent track record in helping develop structural-biology programmes at other universities will catalyse the new initiative.

Huber plans to continue promoting the field by energizing students. "I saw structural biology being born, and now I see it almost matured, providing the basis for our understanding of biology and medicine," he says. ■
Virginia Gewin

RECRUITERS & ACADEMIA

Unhealthy choices

Health benefits are important in countries such as the United States, where there is no national health service. But US postdocs who win funding from sources outside their university often lose benefits and subsidized health insurance, as their institutions don't consider them employees, and the sponsoring institutions don't always provide enough funds to cover benefits.

The National Institutes of Health (NIH) is boosting health-insurance funding for its 7,000 National Research Service award-holders. Good news for NIH-funded postdocs, but the problem goes wider. Benefit options vary greatly among society, government and foundation funders.

According to the 2005 Sigma Xi Postdoc Survey, only a third of postdocs are classified as employees. Still, 97% said their institutions made health benefits available to them, and 82% said families could be included.

But this doesn't reflect a plan's quality, cost or convenience. When Eric Tytell began his NIH fellowship at the University of Maryland, he was told the university's insurance came with so much red tape that if he had any other option, he should take it. He became a dependant on his wife's plan. Adam Breier at the Massachusetts Institute of Technology had to switch

doctors and health plans when he received an NIH fellowship midway through his postdoc; the price of his former plan "was no longer realistic".

Institutions have got better at finding ways to cover non-employee postdocs, says Chris Blagden, a board member of the National Postdoctoral Association. Confronted with postdocs' reluctance to accept NIH fellowships, Case Western Reserve University in Cleveland, Ohio, decided to give the same benefits regardless of funding source, says administrator Rachel Begley. Postdocs are a young, healthy and temporary population, so treating them separately from faculty members and staff means institutions can offer a comprehensive package at a lower cost, says Mary Anne Timmins of the biomedical postdoc programme at the University of Pennsylvania. The University of California has created a separate benefit programme for postdocs on its ten campuses.

But Blagden warns that no solution can fit all institutions, and fixing one problem can cause others. Foreign postdocs who aren't classed as employees can face visa troubles, for example. And postdocs have to look out for themselves, he says. "You have to be aware that the institution doesn't always give you the best choice." ■

Monya Baker

GRADUATE JOURNAL

Roller hockey or science?

What would you be doing if you weren't a scientist? That's what I asked my classmates at the end of our second year of graduate school. By then they'd begun to recognize the difficulty of a career in science. The list of dream alternatives was impressively diverse: four restaurateurs, two concert violinists, one dancer, one fashion designer and a professional roller-hockey player.

What struck me, re-reading this list recently, is how many people's idealized career alternatives were their hobbies. I guess this makes sense. Advice columns are always telling you to make your passion your career. That's what attracted most of us to science in the first place. We were the ones building batteries out of stuff we found in the garage and trying to culture microorganisms from the vegetable drawer. But that childhood passion never included writing grants, rushing to publish or scheduling your life around your experiments.

Then again, I doubt that the would-be restaurateurs were thinking about health inspections or the aspiring singers about contract negotiations. As graduate school has become progressively more stressful, I've increasingly had to remind myself of the passion that brought me here. My hobby — cheese making — helps. I'm not actually tempted to do it professionally, but watching microbes transform milk in my kitchen is still a thrilling display of science. ■

Milan de Vries is a molecular-biology graduate student at the Massachusetts Institute of Technology.

The liquidators

A rose by any other name.

Michael Garrett Farrelly

He keeps telling me in a rolling Kazakh accent to call him Harry Lime, and at 90 miles an hour in the heart of a nuclear wasteland, I'm willing to oblige. I picked up Harry in one of the 'Stans. I can't remember which one, but I know the supreme leader just outlawed citrus for being a "Stalinist holdover". Every time I would ask someone in the 'Stans about the bizarre governments that seemed to bubble up in the region, I'd get a smile and a shrug. "Better than Russia," they'd say, and try to sell me a pirate DVD.

Harry Lime wasn't selling the latest Hollywood hits. He's a tour guide to the exclusion zone. I had to pay him nearly twice the cost of my round-trip business-class ticket for the honour of having him drive me screaming into the Ukraine in his blood-red Chevy SUV. Harry knows all the border guards by name and his bribery with my money is generous to a fault. The last guards on the edge of the exclusion zone don't want a bribe. They just eye us both for a long time and nod us through. My guide with the *Third Man* obsession thinks this is a sign of good fortune.

We slap on three separate rad-readers apiece, all-inclusive to my tour package, thankfully. I like the redundancy more and more as we pass deeper into the zone. There's a lake filled with rusted-out tankers that Harry tells me has enough scrap metal to put Ukraine in the black for the next century. If only the metal wasn't soaked in enough radiation to contaminate all the ground water in Europe.

We hit the town of Prip'yat and we're soaking in 500 microröntgens an hour according to all three of our readers. Harry claims he felt a bump and pulls to a stop to check the tyres and stretch. He warns me not to wander, that pockets of radiation strong enough to make me glow are hiding all around. I walk off towards a high-rise apartment building and keep an eye on all my gauges, feeling like a deep-sea diver back in the day of big metal helmets and hand-powered bellows.

I'd heard the stories about the silence. There was the tour group who all went mad and killed each

other because they couldn't take the utter silence of the region, or the guy who ran screaming towards the reactor itself tearing at his clothes and crying out "Wormwood! Wormwood!" — and a dozen other scary tales. It's silent here, sure, but it's the stillness before something happens. A pregnant pause where you bring your own neuroses and listen to them bounce off the nothing.

Harry honks three times after checking all four tyres and five minutes later we're rolling out. Harry's willing to get me to the edge of the Red Wood forest but no further. He stops and parks about 1,000 yards from where I need to be and turns to me with a smile. I'm shivering so hard I think my heart will stop, and the heater is working just fine.

Harry reaches into the pocket of his battered army coat and pulls out a plastic hourglass, the kind you use for board games. He puts it on the dash and says:

"Ten minutes." I know he means that's my time limit before he pulls off and leaves the crazy yank to rot.

Just outside I slip on the coveralls and seal up the helmet. With dead trees ahead and a taste like iodine in my mouth, I start my run.

The liquidators made this same trip back in 1986. Made it over and over with tonnes of nearly melting slag and bits of reactor shielding. The liquidators got a medal, some extra rations and a miserable death from any number of cancers brought on by exposure. Some of them were up on the roof over the reactor, picking up pieces of core with their bare hands.

It came out here, to Red Wood forest. I'm 200 metres out, according to the GPS. Some suit in Washington once told me over drinks that the CIA should post snipers around the Red Wood forest to protect against terrorists stealing radioactive soil. I'm 15 metres out when I think of the first time someone told me about Chernobyl. A little girl's vision of a smoking crater, complete with Wile E. Coyote standing in the middle with a hand-painted sign reading "Oops".

I kneel on something hard and pray it's not sharp enough to puncture the suit. I slip the vial and mini-trowel out of my external pocket and open it. Borlaug Rose Species 0925, a handful of heavy, dark green pellets. I dig quickly with the trowel, one foot down, and drop the seeds. I'll confess to patting the filled-in hole and saying a little prayer. I'm up and running and Harry's honking the horn as a nervous tic.

Come next spring I'll get Harry, or someone just like him, to bring me back here. There will be a tangle of rosebushes where I was kneeling and the radiation level will be down, at least in the soil. Pretty pink roses, spliced to gobble up radiation, growing in the deadliest place on Earth. Oil slicks turning in floating gardens, chemical spills turned into orchards, that's the goal. Maybe it can all start here where too much ended. ■

Michael Garrett Farrelly is a librarian and writer in Chicago. He can be found online at www.garrettfarrelly.net.



JACEY